TESTING FOR AUTONOMY

Three exploratory case studies of the language testing practices of degree-one teachers of English in the fourth forms of upper-secondary education in the Netherlands

Adri Elsen

2009
Testing for Autonomy: Three exploratory studies of the language testing practices of degree-one teachers of English in the fourth forms of upper-secondary education in the Netherlands
Adri Elsen. Thesis Radboud University Nijmegen, 2009.
With references – With summary in Dutch
ISBN: 978-90-9024137-1

Print: PrintPartners Ipskamp, Enschede

AUTONOMIE IN DE PRAKTIJK GETOETST

Drie explorerende gevalstudies van de toetsingspraktijk van eerstegraads docenten Engels in de vierde klas van het havo en vwo

Een wetenschappelijke proeve op het gebied van de Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van rector magnificus,
prof. mr. S.C.J.J. Kortmann,
volgens het besluit van het College van Decanen
in het openbaar te verdedigen op
dinsdag 12 mei 2009,
om 15.30 uur precies

door

Adrianus Hermanus Maria Elsen

geboren op 12 mei 1956 te Wamel, Nederland

Promotor

Prof. dr. C.L.J. de Bot, Rijksuniversiteit Groningen

Copromotores

Dr. C.J.M. van Esch

Dr. E.V. Schalkwijk, Netherlands Quality Agency, Utrecht

Manuscriptcommissie

Prof. dr. J.W.M. Kroon

Prof. dr M.S. Berns, Purdue University, USA

Dr. P.H.M. van de Ven

TABLE OF CONTENTS

CHAPTER 1: TESTING FOR AUTONOMY: AN INTRODUCTION

The introductory chapter to this study consists of four sections. We will start off with a personal rationale for the present investigation, and refer to two formative experiences that triggered off a lifelong practical and professional interest in assessment and evaluation. We had these experiences in mind, when an appropriate title was required for the endeavour that had just begun. Secondly, we will explain that the title *Testing for Autonomy* is meant to incorporate three meanings. This will be followed by a section on the importance of a teacher's beliefs. Knowledge of teachers' beliefs sensitised us to their potential relevance when we spoke with three practising teachers about central concepts of the subject they taught. In a final section, we will look ahead at what is to come and briefly introduce the chapters of the study. The chapter division illustrates how the study has been structured and how we work towards its findings, reflections and recommendations.

A personal rationale

At the age of eighteen, I was to do my car driving test. The course materials to be studied primarily included theory on traffic rules and regulations. I was not too bothered by that, used as I was to learning things by heart, most of the time to please teachers or parents and occasionally to please myself. The actual test had a theoretical and a practical component. The theory that had to be studied ended with information that really sparked off my interest. There was a section with detailed information on the criteria my future examiner was going to use to assess whether I would be considered capable of driving a car on my own. It also included the actual scoring grids the examiner was to fill in. Never before had I experienced such detailed information on how I was going to be tested. I sensed the information allowed both my driving instructor and me to test my autonomy as a driver in the course of the practical lessons. It gave me hands-on information on what an examiner would ultimately do in a few months' time. The experience was to trigger off a lifelong practical and professional interest in testing.

Then, in my first year of teaching English as a foreign language to Dutch adolescents, I experienced yet another formative episode. I was teaching a class of over thirty 14- to 15-year-olds. In the middle of dealing with a rather dull grammatical exercise, I noticed that one of the sentences referred to American high schools. Looking around, I saw that most of the class were uninterested in the exercise that was being discussed. In an instant I decided to spontaneously tell them what I knew about American high schools and what life in America would be like. As a beginning teacher, I had become aware that learners generally appreciated me talking about subjects that related more to their lives than grammatical exercises generally do. So, I was not that surprised when I got my learners' undivided attention. It was to be a brief excursion from the grammar we were dealing with, and had the effect I was after. After a couple of minutes, one of the youngsters raised her hand and asked me: 'Excuse me, sir. Are we going to be tested on this?'. I distinctly remember being flabbergasted by her question. My intention was not at all to give my learners culture-specific information in order to test them on what they were told. It was just meant to lighten up their form-focused lives a little! Somehow, curious as I was about what would happen, I decided not to tell them what I felt. Instead, my response was: 'Of course, you're going to be tested on this'. What then happened completely took me by surprise. All of the learners, none excluded, started to move and went for their pens and exercise-books in order to copy what I had so far said and written on the blackboard. Early in my career, it gave me the acute sense that tests can be powerful tools that potentially affect whether and in what ways learners learn.

Many years later, I was given the opportunity to elicit and analyse what teachers consider to be effective teaching and testing in the research project that is now to be reported on. That event was to add a third formative episode to the development of

informed views on the roles of assessment and evaluation when learners learn to communicate in another language. It is this very process the present study will report on.

Testing for autonomy

We had the formative experiences above in mind when we were looking for a title to give to the present study. The research theme of our studies was to relate to the teaching and learning of foreign languages. It was broadly meant to address three issues: the learning of pre- and in-service teachers, active and self-regulated learning in school subjects, and didactic procedures that were expected to support that learning in upper secondary education. We arrived at *Testing for Autonomy* for a title. The title incorporates three central meanings.

First, the title refers the researcher's personal test for autonomy in settings where secondary-school learners were expected to learn how to communicate in English. Having experienced first-hand what it was to successfully implement principles of learner autonomy in teacher education (Van Esch, Schalkwijk, Elsen & Setz, 1999), we wondered how alleged principles of autonomy would hold when actually put to the test in secondary education.

The title, therefore, also refers to the ways in which practising teachers would test the autonomy of their adolescent learners. We were to investigate how academically-educated teachers of English would literally test their learners' autonomy in acquiring communicative ability in a year of a massive curricular and didactic reform. In 1999, the regular school subjects of Dutch upper secondary education and some new ones were divided up into four curricular profiles: *science and technology, science and health, economics and society,* and *culture and society*. In addition, the metaphor *Studiehuis* (House of learning) was introduced. *Studiehuis* not only involved a physical change of schools, but also implied didactic procedures that were meant to foster active and self-regulated learning in the adolescent learners. The focus on communicative ability was chosen, because it is an acknowledged domain-specific construct when other languages are taught and learned. Languages are invariably taught, learned and tested with the aim to use them effectively and appropriately as a means of communication.

Third, in addition to our practical interests, our test for autonomy more and more included theoretical explorations of the central constructs of this investigation, which are *learner autonomy*, *communicative language education*, and *foreign language assessment and evaluation*. After we had written first versions of a single theoretical chapter and had started to analyse our teacher data, we soon realised that more theory was required to help us understand and interpret the teacher data in relation to the central constructs. The single theoretical chapter on LA and CLE was soon divided into two. Not much later, a theoretical chapter was added on assessment and evaluation. In the course of this study, our theoretical interests came to be dialectically intertwined with our reflection on and analyses of the teacher data.

Thus, the envisaged *test* for autonomy, soon became a *quest* for autonomy in a process of reflection and interpretation. The originally planned six chapters, developed into twelve. Our aim is to allow the reader to follow the developmental track of this quest for autonomy.

Because we spoke with teachers about their beliefs and interpretations of what they considered to be effective teaching and testing, we will have a brief section on the relevance, importance and nature of a teacher's beliefs. The information helped us to understand and interpret the teacher data and further structured the study.

A teacher's beliefs

From the start of our investigations, we had been sensitised to the fact that speaking with practising teachers about the central concepts of the subjects they teach is likely to elicit a host of beliefs, thoughts, and considerations that include

evaluations, appreciations, doubts, and so on. These beliefs are likely to determine a teacher's practical theories (Handal & Lauvas,1987) or *theories in use* (Argyris & Schön,1974; Kane, Sandretto & Heath, 2002). Handal & Lauvas distinguish three components of teachers' practical theories: *personal experiences*, *transmitted knowledge*, and *core values*.

A teacher first of all brings personal experiences to a situation. These experiences include educational experiences in roles as former pupils, students, as parents, citizens as well as life experiences in general. Past experiences in life, education or work influence who teachers are and what they think and feel when they work as teachers.

A second component distinguished by Handal & Lauvas is transmitted knowledge. It refers to the theoretical and practical knowledge that has been acquired or passed on by relatives, educators, friends, or peers and by media that vary from course materials, newspapers, television, radio, to the World Wide Web. Transmitted knowledge is seen to directly affect teaching practice. According to Handal and Lauvas it includes concepts, categories, theories and commonly held beliefs, which frame teachers' classroom behaviour and shape their notions of what effective language learning and teaching are all about.

As a final component of teachers' practical theories Handal & Lauvas have identified core values. Values and norms originate in what people consider to be just and unjust in life. These values are culturally determined and may vary in quantity, specificity and in the ways in which they are interpreted. Values and norms may relate to ethics, philosophy, politics, religion, education and so on. They may follow mainstream cultural points of view or can fundamentally differ from them. Sometimes they are global, at other times very concrete. Core values identify us as human beings in all of the social roles we have.

Zeichner & Liston (1996) have stated that teachers' practical theories are not the only determining factors of classroom practice. A crucial factor is the social context in which the teachers work. Examples of such contextual factors are governmental and school policies, the typical ways in which a school is organised, the ways in which teachers are facilitated to develop as professionals, or the effects brought about by the use of particular course materials. Therefore, both a teacher's practical theories as well as external factors seem to influence the way in which teachers respond to the opportunities and constraints that surround them. It is suggested that teachers critically reflect on their practical theories as well as on the constraints and opportunities caused by factors outside immediate control. Zeichner & Liston (1996) state that a reflective teacher (a) examines, frames and attempts to solve dilemmas of classroom practice; (b) is aware of and questions the assumptions and values he/she brings to teaching; (c) is attentive to the institutional and cultural contexts in which he/she teaches; (d) participates in curriculum development and school change efforts; and finally (e) takes responsibility for his/her own professional development.

A host of additional literature and research on teachers' cognitions in relation to their classroom practices further sensitised us to the importance of teachers' beliefs (for overviews see Pajares, 1992; Mathijsen, 2006: 164-166). We were also interested to learn that research on beliefs can have a more domain-specific focus and direct attention from what a teacher believes to what their learners believe. Learner beliefs about learner autonomy have been studied in several research projects. Examples of such studies are Cotterall (1995) and Crabbe, Hoffmann and Cotterall (2001).

Cotterall (1995) argues for the necessity to probe and assess the learners' readiness for the changes in behaviour and beliefs which autonomy implies, before any intervention of autonomy-oriented approaches to language learning is made. Her study presents data on learner beliefs collected in a study with 139 adult ESL learners in the summer of 1992-1993. The overall aim was to see if subjects' responses to a questionnaire revealed any particular cluster of beliefs. Using a factor

analysis, six factors were obtained: (1) Role of the teacher, (2) Role of feedback, (3) Learner independence, (4) Learner confidence in study ability, (5) Experience of language learning (i.e. experiencing being successful in language learning) and (6) Approach to studying. Her conclusion was that these learner beliefs would affect (and sometimes inhibit) learners' receptiveness to the ideas and activities presented in the language class. In her own words: “Learners’ beliefs about language learning will profoundly affect their approach to language learning” (Cotterall, 1995: 202). She claims that an awareness of the role of cognitive and affective variables in language learning, of how language functions and of how strategies influence learning can enhance learners’ quality of thinking and the ways in which they engage in tasks. Autonomous learners, she maintains, draw on their experiences of working on tasks, using strategies and solving problems to deepen their understanding of the target language system. “By exploring these beliefs, learners and teachers can hope to construct a shared understanding of the language learning process and of the part they play in it. This awareness is an essential foundation of learner autonomy.” (Cotterall, 1995: 203).

Research by Crabbe, Hoffmann and Cotterall (2001) further suggests that a teacher’s awareness of learner beliefs is a decisive factor in successfully implementing learner autonomy. Analysing the discourse of interviews between three learners of a second language at university level and their tutors, the researchers have explored the problems experienced by these learners in immediate and long-term goal setting and in expressing their beliefs about language learning. In their conclusions, they state that goals occupy a central position in the learning process and that “making the goals explicit seems a useful basis for any subsequent discussion of strategic behaviour that might serve those goals”. They also claim that “the immediate effectiveness of the session could be measured by how well the learners represent the problem, how committed they are to specific goals, and how aware they are of their beliefs about language learning” (2001: 14).

Although we fully acknowledge the importance of learner beliefs, the present study focuses on what teachers believe and on how they perceive what their learners think and feel, with their actual tests used as interfaces between what teachers want and what learners allegedly do. How teachers interpret three crucial constructs and the ways in which their beliefs and interpretations affect how they teach and test and how their adolescent learners are expected to learn is the central interest of this study.

In the final section of our introduction, we will look ahead to the chapters to come.

What is to come

In chapter 2: *The exploratory multiple-case study*, we will provide details on the investigation. We have already discussed a personal rationale in this introductory chapter. In the next chapter, we will adopt a more objective stance. First, we will go into its rationale, objectives, questions, and method. Next, we will provide information on how the teacher respondents have been selected, the five stages in which the study is to be carried out, and the ways in which the investigation will be documented, analysed and reported on.

Chapter 3: *Learner autonomy as a pedagogical construct* is the first of our theoretical chapters on the constructs central to this investigation. Here, we will address three questions: what learner autonomy is, why it can be seen as a viable pedagogical construct and goal, and what the relationship is between learner motivation and autonomy.

In chapter 4 we will address the construct of *Communicative competence in foreign language education* and explore some of its backgrounds. We will start with a concise methodological history of second and foreign language learning and teaching. Next, we will turn to methodologies that aim at meaningful communication and a specification of the knowledge and skills required to communicate correctly,

efficiently and appropriately. The section will be followed by a discussion of influential models of communicative competence and three interpretations of the construct, characterised by a focus on form, meaning, or context. We will end our second theoretical exploration with some of the more recent research insights in communicative ability.

Chapter 5 will deal with the construct of *Foreign language assessment and evaluation*. We will first of all stress the importance of clearly distinguishing between the notions testing, assessment and evaluation. Secondly, we will review four trends in the field of language testing, subsequently the *pre-scientific*, *psychometric-structuralist*, *integrative-sociolinguistic*, and *critical-dynamic* trends. In a third section we are going to survey three subject domains that have helped to establish professional standards in the field of language testing. We will first discuss tests in terms of test purpose and test types. Then we will go into the essential measurement qualities of reliability and validity. Finally, we will deal with Bachman & Palmer's (1990) definition of test usefulness as the sum of six interrelated qualities, namely *reliability*, *construct validity*, *authenticity*, *interactiveness*, *impact* and *practicality*.

In chapter 6: *A context of innovation* we will temporarily interrupt the momentum of our study to ponder over the context of secondary and teacher education in the Netherlands amidst turbulent curricular and didactic reform. The next four chapters will consist of reports on the teacher data that were gathered in the year of data collection.

Chapter 7: *Three stories to tell* is a predominantly narrative chapter, in which the three respondent teachers, their core beliefs and their construct interpretations will be introduced. The chapter is based on the interview data we gathered before the start of the school year.

Chapters 8, 9 and 10 will report on the teacher data collected in the course of the school year, when three sample written English tests were selected by and discussed with the three informants in the light of the research questions. Chapter 8 focuses on *Joy, the budding professional*, chapter 9 on *Mark, the literary master*, and chapter 10 on *Pete, the project man*.

In chapter 11: *Cross-case analyses*, the data of this study on LA, communicative language education, and foreign language assessment and evaluation will be reduced, compared and contrasted by way of cross-case analyses, and subsequently related to the contents of the three theoretical chapters.

Finally, chapter 12 *Autonomy tested*: a discussion will first look back on the seven research questions. We will do so in an effort to establish which questions have already been addressed in the previous chapters, and determine which questions remain to be answered. Secondly, in a section called Theory and the teacher data, we will look into the question in how far the research findings compare and contrast with the theoretical insights discussed in the chapters on learner autonomy, communicative language education, and foreign language assessment and evaluation. Thirdly, in the section named A context of innovation and reform, we will take up the question in how far the second-phase reform has been conducive to fostering learner autonomy, enhancing communicative language ability, and promoting effective assessment and evaluation. Fourthly, we will arrive at Recommendations for pre- and in-service teacher education. In this penultimate part, we will explore the question what can be learned from the investigation in view of further research and educational programmes on how to create positive washback of assessment and evaluation practices on classroom teaching and learning in settings where learners learn to communicate in another language. Finally, we will go into what we see as crucial shortcomings of our qualitative and interpretative investigation. These shortcomings require us to be modest, despite all the insights, understandings and critical concepts the investigation has helped us arrive at.

CHAPTER 2: THE EXPLORATORY MULTIPLE-CASE STUDY

2.1 Introduction

In this chapter we will first present the rationale of the study and explain why *Testing for Autonomy* focuses on the three constructs: *communicative language education, learner autonomy*, and *foreign language assessment and evaluation*. Next, we will present the research objectives and questions we attempt to answer. The objectives and questions favoured the use of a case study approach and design. We will justify our choice for an embedded multiple-case design in a next paragraph. In the paragraphs that follow, we will discuss details of the multiple-case study, such as the initial theoretical framework, the selection of the three teacher respondents, the various research stages and the specific questions asked. In three closing paragraphs we will mention how the case study evidence was documented, analysed, and presented.

2.2 Rationale

In the Netherlands, communicative approaches led to the design of new foreign language course materials in the 1980s. Language functions, communicative situations, and oral proficiency gradually became more prominent in classroom tasks and assignments. The 1993 targets of basic secondary education, i.e. the first two or three forms of the various types of secondary education, were a first attempt to specify communicative language ability in more detail.

Communicative syllabuses were also promoted in the 1999 reform of foreign languages in upper secondary education, i.e. forms 4 and 5 of *havo*-education, and forms 4-6 of *vwo*-education. This massive curricular and didactic reform is generally referred to as *second phase.* Second-phase didactic approaches aimed to foster more learner autonomy in the learners than had so far been the case. More than ever before, foreign language learners were now expected to acquire and develop attitudes and skills essential to self-regulation and control in foreign language learning, even though these attitudes and skills were not specified in targets or benchmarks.

The reforms of Dutch junior and upper-secondary education were introduced top-down, with little or no consultation of the field. Schools and teachers were given some time to experiment, but were obliged to work along the new guidelines from August 1993 (basic secondary education) and August 1999 (second phase). The second-phase reform meant that foreign language teachers in the upper forms of secondary education were given fewer classroom lessons, as the adolescent learners were expected to partly work and learn on their own. On top of this, most teachers were given more forms or grades to teach.

The reform also affected testing practice. All of the informal tests in secondary education had to be protocolled and planned at the beginning of the school year in so-called PTAs (*Programmes for Assessment and Summative Evaluation*). More than ever before, the teachers were expected to think about test content, format and timing. The context of the 1999 *second phase* reform was the backdrop of the present research.

As before, teachers would teach and test. Yet, little could be hypothesized about how the second-phase reform in the Netherlands would affect the teachers' informal assessment and evaluation practices. In addition, research has often shown how little

we know about what actually goes on in a foreign language classroom. Relatively few empirical studies concentrate on the actual teaching and learning processes in the classroom. An exploration of the complex relationship between teaching and testing was an important impetus for the present study.

The fact that language tests can be powerful and influence teaching and learning processes is acknowledged by a lot of teachers and learners. The impact of language tests on teaching and learning is called *washback* or *backwash*. These effects of tests can either be beneficial or harmful to teaching and learning, resulting in *positive* or *negative washback* (Morrow, 1986; Hughes, 1988; Frederiksen & Collins, 1989; Alderson & Wall, 1993; Wall & Alderson, 1993). Most teachers know that language tests may function as learning incentives and are familiar with questions like “Madam/Sir, are we going to be tested on the stuff we’re dealing with?”. Teachers are also familiar with the effects of national examinations on the foreign language curriculum, and are often critical of the washback of examinations on classroom teaching. Alderson & Wall studied the so-called washback hypothesis in greater detail. At its most general, the hypothesis claims that *a test will influence teaching and learning.* To illustrate the complexity of the construct, Alderson & Wall extend the general assumption with related washback hypotheses.

Some possible Washback Hypotheses

(1) *A test will influence teaching*.

This is the Washback Hypothesis at its most general. However, a second partly different hypothesis follows by implication from the first one, on the assumption that teaching and learning are related, but not identical:

(2) *A test will influence learning*.

Since it is possible, at least in principle, to separate the content of teaching from its methodology, then we need to distinguish the influence of a test on the content of the teaching from its influence on the methodology. Thus:

(3) *A test will influence **what** teachers teach;* and
(4) *A test will influence **how** teachers teach;* and therefore, by extension from (2) above:
(5) *A test will influence **what** learners learn;* and
(6) *A test will influence **how** learners learn.*

However, perhaps we need to be somewhat more precise about teaching and learning, in order to consider how quickly and in what order teachers teach and learners learn. Hence:

(7) *A test will influence the **rate*** and ***sequence*** of teaching; and
(8) *A test will influence the **rate*** and ***sequence*** of learning.

Similarly, we may wish to consider explicitly both the quality and the quantity of teaching and learning:

(9) *A test will influence the **degree** and **depth** of teaching;* and
(10) *A test will influence the **degree** and **depth** of learning.*

If washback relates to attitudes as well as to behaviours, then:

(11) *A test will influence attitudes to the content, method etc. of teaching and learning.*

In the above, however, no consideration has been given to the nature of the test, or to the uses to which scores will be put. Yet, it seems not unreasonable to hypothesize:

(12) *Tests that have important consequences will have washback;* and conversely

(13) *Tests that do not have important consequences will have no washback.*

It may be the case that:

(14) *Tests will have washback on **all** learners and teachers.*

However, given what we know about difference among people, it is surely likely that:

(15) *Tests will have washback effects for **some** learners and **some** teachers, but **not** for others.*

(Alderson & Wall, 1993b:120/1)

Alderson & Wall carried out a two-year longitudinal study of the impact of a new English examination in Sri Lanka on teaching. Their research has shown that washback is a complicated construct, which in the Sri Lankan case influenced the contents of teaching, but did not affect how the contents were taught. They recommended further exploration of the often oversimplified relationship between tests and teaching in classroom practice, particularly in times of curricular innovation. In this study, their recommendation is taken up by considering teachers' beliefs about washback and the relationship between teaching and testing as sensitising concepts.

Yet, there is no teaching or testing without content. This content is largely determined by three dominant constructs that stand out in the reform of foreign language secondary education in the Netherlands: *learner autonomy, communicative language education,* and *language assessment and evaluation*. We decided to investigate the three notions in the micro-context of the English language classroom.

What the three constructs have in common is a deceptive transparency. A lot of foreign language teachers would readily acknowledge that the three constructs are relevant to their classroom practices. After all, teachers ultimately wish their learners to become competent and proficient users of the target language who manage to communicate in the language correctly, effectively, and appropriately. In addition, many a teacher feels that learner autonomy is a construct hard to find fault with. It will be difficult to find foreign language teachers who do not at least hope that the knowledge, skills and understandings of their teaching will ultimately be put to good use by their learners. Assessment and evaluation is yet another construct easily taken for granted. Over the years, teachers tend to arrive at approaches to language testing they more or less consider practical and effective in view of their intended purposes. These approaches result in tests that teachers administer and mark. Following the washback hypothesis, these tests may be important instruments that help teachers as well as learners to assess and evaluate progress. A focus on the

actual tests that teachers construct and use and the results of these tests is particularly relevant in view of the intended improvement of the learners' knowledge and abilities in the curricular and didactic reform of upper secondary education.

So the meanings of the three constructs of this investigation tend to be seen as self-evident, with autonomy and communicative ability considered as logical goals and language tests accepted as assessment procedures. This study very much focuses on the ways in which the three main constructs can be interpreted. We will do so by investigating the constructs from a theoretical point of view and gather classroom data. Thus, theoretical and practical interpretations of learner autonomy, communicative language education, and foreign language assessment and evaluation will be compared and contrasted.

This study has been based on monitoring three good-practice teachers of English and their fourth-form adolescent learners in the school year 1999-2000. The forms taught by the teachers represent the three types of upper secondary education in the Netherlands: *havo (senior secondary education), atheneum (modern grammar school)* and *gymnasium (classical grammar school)*.

2.3 Research objectives

The study is part of a research programme on the professional development of secondary school teachers carried out by ILS, Graduate School of Education, Radboud University Nijmegen, the Netherlands. This means that the research objectives of the investigation are related to the main objectives of the ILS research programme. The key theme of the programme is teacher professionalism studied in classrooms as well as in the educational institutions in which they are employed. The programme consisted of four related project areas:

1. learning of (future) teachers;
2. active and self-regulated learning in school subjects;
3. teaching and learning of foreign languages;
4. learner autonomy didactics in upper secondary education.

The ILS research projects aim to generate theoretical and empirical knowledge related to teacher professionalism. Theory is linked to practice and practice to theory to arrive at a better understanding of the phenomena under investigation. Empirical data are often based on series of dialogues with teachers and analysis of teacher discourse. Links between academic knowledge and school practices have been explored in a variety of studies carried out under the auspices of ILS (Theunissen, 1996,Van Schalkwijk, 1998, Van Amelsvoort, 1999, Luttenberg, 1999, Kamp, 2000, Engelen, 2002, Smits, 2003, Marijnissen, 2003, Van Veen, 2003, Reis, 2005, Hermans-Nymark, 2006). The present study is primarily aimed at teacher professionalism in relation to the three constructs under investigation in the context of second phase reform.

The investigation focuses on English as a secondary school subject and, as such, is part of project area 3, that is the teaching and learning of foreign languages. Despite its subject-specific focus, the study is related to the other three project areas. It is first of all related to the learning of (future) teachers. By concentrating on the key theme of assessment and evaluation, the study aims to arrive at recommendations for viable training programmes for (future) teachers on how to create positive washback of a teacher's assessment and evaluation practice on *what* the learners learn, *how* they go about this, and *why* they do so when they learn to communicate in a foreign language. Secondly, assessment and evaluation is an essential component

of any teaching-learning process and is of particular interest when active and self-regulated learning is a main didactic target. Exploration of the role of assessment and evaluation in active and self-regulated learning is an issue directly addressed in the present study. Finally, the didactic reform of upper secondary education is a key contextual factor during the study. This context of reform enables us to consider in how far the innovations are conducive to teachers' interpretations of learner autonomy and to the ways in which they are operationalised in three classroom practices.

2.4 Research questions

The study attempts to answer three questions in the context of the second-phase reform of upper secondary education in the Netherlands.

1. What are the beliefs of degree-one teachers of English with regard to *learner autonomy*, *communicative language education,* and the role of *assessment and evaluation*?
2. How are their beliefs reflected in their assessment and evaluation practices?
3. What recommendations can be made to enhance learning environments in which tests have beneficial washback effects on the ways in which learners learn to communicate in English?

The main research questions above have been specified in seven related questions.

1. What can we learn from studying theory on *learner autonomy, communicative language education,* and *foreign language assessment and evaluation*?
2. How do teachers of English define and specify *learner autonomy, communicative language education,* and *effective foreign language testing*?
3. What beliefs, experiences and arguments underlie the design and administration of three sample informal tests constructed by the teachers themselves?
4. How do teachers interpret the test results of the sample tests in relation to their views, opinions and beliefs of communicative language education and learner autonomy?
5. How do the research findings compare and contrast with theoretical insights in *learner autonomy, communicative language education,* and *foreign language assessment and evaluation*?
6. In how far has the second-phase reform been conducive to fostering learner autonomy, enhancing communicative language ability, and promoting effective assessment and evaluation?
7. What can be learned from the investigation in view of further research and training programmes on how to create positive washback of assessment and evaluation practices on classroom teaching and learning?

2.5 Method

With a focus on three good-practice teachers of English and multiple units of analysis, an embedded multiple-case design best suited the purposes of the

investigation. The exploratory nature of the investigation, the complexity of the constructs, the contextual conditions and its reliance on multiple sources of evidence justified this type of qualitative research design. Data was gathered on the basis of four related interviews, one at the beginning and three in the course of the school year under investigation. The units of analysis consisted of teacher interviews at the beginning of the school year, teacher interviews in the course of the school year, classroom observations, and nine written language tests selected and constructed, administered and marked by the teachers. All of the interviews were protocolled by way of interview guides. The design was meant to result in converging evidence of convincing reliability and internal validity in the course of the investigation.

Before the data was collected, publications relevant to the three constructs under investigation were studied. The constructs were studied from perspectives derived from educational sociology, philosophy, educational psychology and applied linguistics. Next to possible washback hypotheses, the initial framework provided searchlight theories that helped us design the multiple-case study and interpret its first results. The framework was extended and revised after each round of data analysis, and finally resulted in three theoretical chapters on the constructs under discussion.

Teacher data was collected in four subsequent stages throughout 1999 – 2000. In *stage O* the initial situations of each of the three teachers were determined before the school year had started by means of semi-structured free-attitude interviews. The three stages that followed focused on the three sample tests constructed, selected and administered by the three teacher respondents in the course of the school year 1999-2000. The levels of intervention of *stages 0, I, II and III* were controlled and limited to asking semi-structured, open-ended questions geared to collecting information relevant to the questions the study was meant to answer. All of the interviews were based on piloted versions of *interview guides* that were offered to the respondents for study at the beginning of each interview. A limited number of classroom observations were carried out to check in how far the teacher rhetoric displayed in the interviews corresponded with actual classroom behaviour.

Each of the research stages was set out in protocols and documented. Special care had been taken to ensure appropriate levels of construct validity and external validity. Using four related interview guides helped to establish a chain of converging evidence. This approach helped to enhance the construct validity of the study. The external validity pertains to the generalizability of the findings of the investigation. Whereas quantitative research generally relies on statistical generalization, the external validity of case studies is often achieved by *analytical generalization* (Yin, 1994: 36), also called *analytic generalizability* (Firestone, 1993). In the case of analytical generalization, the investigator attempts to generalize a particular set of results to some broader model or theory.

> *A common complaint about case studies is that it is difficult to generalize from one case to another. Thus analysts fall into the trap of trying to select a "representative" case or set of cases. Yet, no set of cases, no matter how large, is likely to deal satisfactorily with the complaint.*
> *The problem lies in the very notion of generalizing to other case studies. Instead, an analyst should try to generalize findings to "theory", analogous to the way a scientist generalizes from experimental results to theory. (Note that the scientist does not attempt to select "representative" experiments.* (Yin, 1994: 37)

The present research strives to generalize the teacher data to theoretical insights with regard to *learner autonomy, communicative language education, foreign*

language assessment and evaluation. No claims whatsoever will be made to having come up with representative cases that can be generalized to other populations and settings.

The interviews were coded and analysed by using KWALITAN, a programme developed by Peters and Westers of the Radboud University Nijmegen. (www.kwalitan.net). The software has been based on the *Grounded Theory Approach* (Glaser & Strauss, 1967; Strauss & Corbin, 1990; Miles & Huberman, 1993; Peters & Wester, 1995) and was helpful in documenting and analysing the chains of evidence.

2.6 The Multiple Case Study

First, a project proposal was written in which the issues to be studied were defined and the case study design was developed. This was followed by a study of secondary reading related to the three main constructs, which led to the initial framework mentioned before. In the course of the investigation, this first framework was revised and extended to the present theoretical chapters 3, 4 and 5, respectively on learner autonomy, communicative language education and assessment and evaluation. Care was taken to carefully document each and every stage of the research. This was done from the start by writing a research hand-out that was used as a case study protocol.

2.6.1 The selection of teacher respondents

Two teachers of English, Mark and Pete, were selected from a group of fifteen teachers that seemed to represent good practice in English language teaching and testing as we had preliminarily defined the term. All of the fifteen teachers worked or had worked as school supervisors of student teachers. The third teacher, Joy, had recently graduated at the graduate school of education as a degree-one teacher of English.

We again stress that no claim will be made to having come up with representative cases, as is typical of positivist quantitative research paradigms. The three teachers were selected on the assumption they would render a wide variety of data within the usual constraints of a time-consuming multiple-case study.

The three respondents were employed at two different secondary schools and taught fourth forms that represented the three levels of upper secondary education in the Netherlands: *havo*, *atheneum*, and *gymnasium*.

Joy, Mark, and ***Pete*** were felt to reasonably represent the good practice criteria that had been formulated at the start of the study. We will try to do justice to their attitudes, beliefs and experiences as expressed in the first interviews in chapter 7 entitled *Three stories to tell.* The good practice criteria are listed below.

Excellent class management

The teachers would have to be good classroom managers, known for their clear and often challenging instructions, efficiency in organizing their lessons, and skills in preventing disciplinary problems before they occurred. This criterion was considered important.

Beforehand we wished to rule out any problems related to classroom management and lack of clarity of instruction, which would interfere with our exploration of assessment and evaluation practice.

Used to developing and constructing their own course materials and tests

It was assumed that teachers who hardly ever constructed their own course materials and almost always used ready-made tests with little adaptations over the years were less likely to render relevant research data.

Generally appreciated as English teachers by their school management and the learners.

In the course of school visits, the researcher often informally speaks with school managers and pupils, and observes teachers and learners in their everyday practices. The information was used in the selection of the teacher respondents.

Known to be seriously interested in the individual learner, that is in their initial situations and development as individual learners.

We were looking for teachers seriously interested in the individual learners and their learning processes, who could be expected to provide details on most of the pupils in a class at any given time.

Known to be precise and accurate in making arrangements related to school affairs.

Visits had to be planned for interviews and observations. It was felt that some accuracy in this area would be helpful, and would possibly be indicative of arrangements the teachers usually made with their learners.

Aiming for high levels of knowledge of and skills in English geared at communication.

The teachers would have to be interested in foreign language learning and teaching that transcends the reproduction of grammar rules and vocabulary. They would have to be interested in improving their learners' communicative language skills.

Ability to clearly voice what they believe to be effective language teaching and learning.

Exploratory research by Setz (1999, unpublished) had shown that some teachers had great difficulty in voicing complex constructs related to fostering autonomy in foreign language learning and teaching. The teachers in this study would be asked to define effective language teaching, learning and testing and by doing so voice how they define communicative language education, learner autonomy in language learning and effective language assessment and evaluation. Because the study aimed to look for opportunities of test washback rather than illustrate its limitations, it was considered necessary to relate the respondents' teaching and testing practices to clearly-voiced beliefs.

Ability to clearly mention and explain purpose, content and assessment criteria of a self-constructed language test.

Foreign language assessment and evaluation is related to purpose, content and assessment criteria. The teacher respondents' ability to supply information on purpose, content and criteria was considered important. It would give both the respondent and the researcher the opportunity to interpret test results in a valid way and to determine what can or what cannot be done to stimulate positive washback of the assessment procedure for any future learning.

2.6.2 The stages of data collection

To realise the overall objectives and attempt to answer the questions of *Testing for Autonomy,* a number of issues were addressed before (**stage 00**) at the beginning (**stage 0**) and in the course of the multiple-case study (**stages I, II & III**).

Initial theoretical framework (Stage 00)

Initially, Elsen & Setz set out to define and clarify the constructs of the studies in which we were involved. An essential research question was answered in this first theoretical framework: In how far does a study of concepts of self-determination in societies (educational sociology), theories on learning and motivation (educational psychology) and developments in the learning and teaching of foreign languages (applied linguistics) help to investigate and interpret the present interest in learning to learn how to communicate in a foreign language?

The initial framework was extended and revised over the part-time years it took to analyse the data and finish the study, resulting in three theoretical chapters: chapter 3 *Learner autonomy as a pedagogical construct*, chapter 4: *Communicative language education*, and chapter 5 *Foreign language assessment and evaluation.* The sixth chapter, entitled *A context of innovation*, emphasises the context of education in the Netherlands and presents details of the second-phase reform.

Initial situations of the teachers involved (Stage 0)

As an example the semi-structured open questions out of the first interview with the teachers are presented. The questions were set out in a protocol and controlled by means of interview guides.

- How has your language testing practice developed from the moment you tested learners for the first time?
- What do you believe to be effective English language teaching?
- How would you define a *good informal written test* and what are its characteristics?
- What opportunities and/or drawbacks of your teaching and testing practice can you mention?
- What *knowledge, insight* and *skills* do you consider relevant for your learners to learn?
- How would you define *learner autonomy, communicative language education,* and *effective foreign language testing*?
- What do you feel are the roles of test items and language tests when learners learn to communicate in English?
- What important *decisions, arrangements* or *plans* can you mention related to your language testing practice in the fourth forms of Dutch upper-secondary education (*havo, atheneum* and *gymnasium*)?

Stages I, II & III

Additional research questions related to the *construction* and *use* of three informal written tests were singled out by each of the teachers. Central to the three stages is exploration of the roles these tests have in advancing the learners' communicative ability in English and in fostering their autonomy.

- Why did you select the test under discussion as a sample test?
- What knowledge, insight and skills are tested?
- How was the test constructed and used?
- How did you prepare the learners for the test?
- How do you think the learners prepare for the test?
- What are the test conditions?
- How are the test results assessed and/or evaluated?
- What have you concluded or decided after interpretation of the test results?
- What has led to these conclusions or decisions?
- What feedback on the test results is offered to the learners?
- How does the test affect the learners' listening, speaking, reading and/or writing skills in English?
- In how far does the test affect how learners regulate their own learning?

2.6.3 Documenting the case study evidence

The case study database consists of twelve teacher interviews, nine sample tests, and four classroom observations. Additional data provided by the teachers consists of diaries or notebooks, extra tests and/or materials they had provided related to the tests they had selected.

This results in nine types of documents that were used for data analysis and report:

- interview guides functioning as interview protocols. (The two types of interview guides are presented as appendices 4 and 5);
- transcripts of the interviews recorded on minidisk;
- case study notes made during and after an interview or observation;
- English translations of the Dutch transcripts carried out by a native speaker of American English raised in the Netherlands and revised by the researcher where necessary;
- teachers' accordance with the transcripts of the first interview before the second interview was held;
- notes taken during classroom observations;
- copies of the sample tests;
- copies of the teachers' diaries and marks assigned to the learners;
- opies of extra materials offered to the learners in relation to the tests that had been discussed.

2.6.4 Analysing the case study evidence

The data of *testing for Autonomy* was analysed in the following stages:

- assignment of open codes and four categories of memos to the teacher interviews by using KWALITAN;
- data reduction by way of data matrices. The matrices led to the selection of relevant interview segments, the reformulation of codes, and the addition and/or revision of memos if required;
- analysis of the classroom observations to check for any discrepancies between teacher rhetoric and their classroom behaviour;
- descriptive analyses of the tests;
- identification of key themes per teacher respondent related to learner autonomy, communicative language education and assessment and evaluation;
- cross-case analyses of the key themes, subdivided in the three constructs under investigation;
- interpretations of the cross-case analyses by analytically generalising them to the theoretical insights discussed in the theoretical chapters;
- interpretations of the data in view of the context of second-phase reform.

2.6.5 Reporting on the case study evidence

We have chosen to report on the teacher data of the study in four chapters with frequent quotes from the interviews. This has been done for two reasons. First, it is felt that narratives based on direct quotes will bring the three respondent teachers to life in a more convincing way than would be the case with tables and isolated utterances. Secondly, we feel the choice for narratives helps to present the data in a way open to the reader's interpretation and validation.

Chapter 7 *Three Stories to Tell,* presents the data from the first interviews held before the actual school year had started. Chapter 8, 9 & 10 document the follow-up interviews on the three tests that were held with the three teachers in the course of the school year. Chapter 8 is on *Joy, the budding professional*, chapter 9 on *Mark, the literary* master, and 10 on *Pete, the project man.* In chapter 11, entitled *Cross-*

case analyses, the teachers' core beliefs and construct definitions are first compared and contrasted and subsequently related to the theoretical insights discussed in the three theoretical chapters. Chapter 12 *Autonomy tested: A discussion* attempts to do what it promises.

2.7 Summary

In this chapter we have first offered the rationale and relevance of the present study. We mentioned the backdrop of the massive 1999 curricular and didactic reform of upper secondary education in the Netherlands, commonly referred to as *second phase*. The reform explicitly focused on learner autonomy, by actively involving the learners and developing their abilities for self-regulation. The second-phase reform had been preceded by a 1993 reform of basic secondary education, which could be referred to as a kind of 'first phase'. For the foreign languages, the emphasis of this reform had been on the specification of communicative language ability and enhancement of communicative language education. The 1999 reform also affected testing practice. With the so-called washback hypotheses as sensitising concepts in the back of our minds, we assumed that in the first year of the 1999 reform, teachers would continue doing what they had been doing for years, that is they would teach and test. This resulted in attention to the three central constructs of this study: *learner autonomy*, *communicative language education*, and *foreign language assessment and evaluation*.

Next, we dealt with the research objectives and the three general questions, and the seven more specific questions the present study attempts to answer. We claimed that projects carried out under the auspices of ILS (Graduate School of Education, Radboud University Nijmegen) aim to generate theoretical and empirical knowledge related to teacher professionalism. This is done by linking theory to practice and practice to theory with the aim to arrive at a better understanding of the phenomena under investigation. The three main questions we attempt to answer are:

- What are the beliefs of degree-one teachers of English with regard to *learner autonomy*, *communicative language education,* and the role of *assessment and evaluation*?
- How are their beliefs reflected in their assessment and evaluation practices?
- What recommendations can be made to enhance learning environments in which tests have beneficial washback effects on the ways in which learners learn to communicate in English?

Finally we provided details on the multiple-case study. We justified our choice for an embedded multiple-case design. Next, we discussed on what criteria the three respondents teachers Joy, Mark and Pete had been selected. Then we provided more detailed information on how the data was collected in four consecutive stages 0, I, II, and III. Finally, we dealt with how the data was documented, analysed and reported on.

In the next three chapters we will concentrate on the constructs central to this investigation.

CHAPTER 3: LEARNER AUTONOMY AS A PEDAGOGICAL CONSTRUCT

3.1 Introduction

In three consecutive chapters we are going to explore the central constructs of the present study. In this first theoretical chapter, we will consider learner autonomy (LA) as a pedagogical construct and goal. The chapter addresses three questions: ***what*** LA is, ***why*** it can be seen as a viable pedagogical construct and goal, and ***what*** the relationship is between learner motivation and autonomy.

In the section on what LA is, we aim to arrive at essential parameters and conditions of autonomy. These parameters will be derived from comparing and contrasting a number of definitions that have been influential in studies on LA. The parameters and conditions we aim for will be compared and contrasted with the definitions of LA given by the three respondent teachers in the course of this investigation in chapter 11.

Why learner autonomy is a significant pedagogical construct is shown by considering LA more closely from four theoretical perspectives. Our first perspective is *socio-historical*. We will first discuss the individual cases of Socrates and Comenius in an effort to show that there have always been successful teachers holding liberal philosophies of education who explicitly aimed at increasing the autonomy of their learners. Next, we move on to a discussion of four culture orientations that have alternately dominated Western societies. We end our socio-historical exploration with a discussion of the requirements of today's knowledge societies, with their need for knowledge, which is growing rapidly and changing fast. All in all, the socio-historical perspective offers three reasons why LA is a viable pedagogical construct and goal.

Our second theoretical perspective is *philosophical*. We will subsequently discuss four prevalent accounts of human agency and autonomy, i.e. *coherentist* approaches, which stress the importance of motives and motivation, *reasons-responsive* approaches, which focus on the importance of individual choice from a wide variety of motives, *responsiveness-to-reasoning* approaches, in which the quality of reasoning of a particular motive to act is highlighted, and finally *incompatibilist* approaches, which claim that people can never be completely autonomous, because there is always interdependence of the self with others or interference with external factors. The *reasons-responsive* and the *responsiveness-to-reasoning* approaches provide a fourth justification why LA is a relevant educational goal.

Our third theoretical perspective is *political*. Autonomy can be considered from a political point of view. At issue is in how far a learner is allowed control over his/her own learning and over what is learned? After all, it is often the state that decides on school curricula, the quantity and level of school exams, or on any standardised national examinations. A political version of autonomy, therefore, focuses on the creation of structural conditions that enable learners to not only control their own learning, but also partly determine the conditions and content of what they learn. Political perspectives on autonomy are closely linked to critical theory, critical pedagogy, and postmethod pedagogy. A political perspective on autonomy strongly invites us to make ethical choices in the kinds and levels of autonomy that will be given to adolescent learners. It is a fifth justification of LA as a pedagogical construct.

A *psychological* perspective invites us to discuss important theories of knowledge and learning that each help to justify a focus on LA in formal education. We have selected three influential views of knowledge and learning for discussion, i.e. a *behaviourist* view, in which desired behaviour is reinforced by repetition and positive feedback; a *humanistic* view, in which personal competence, self-respect and intrinsic and integrative motivation are stressed; and finally a *cognitive* view of human

learning, with more detailed attention to constructivism, Piaget's genetic epistemology and Vygotsky's sociocultural theory. The humanistic and cognitive views of learning in particular are in strong support of a focus on learner autonomy in pedagogical contexts and offer suggestions of how learner autonomy can be fostered in formal educational settings.

Because the definitions of autonomy and the theories of mind we are to discuss assign crucial roles to learner motivation, we will explore the relationship between autonomy and motivation in more detail in the one but last section.

We will end our first theoretical chapter[1] with a summary of the sections on the *what* and *why* of LA.

This chapter is to offer a frame of reference that may help us interpret and understand the LA-related findings of this study and their pedagogical implications for the implementation in upper-secondary classes of English in the Netherlands.

3.2 What is learner autonomy?

Defining *learner autonomy* is no easy matter. When foreign language learners, teachers, teacher educators or researchers are asked what learner autonomy is, one is likely to get a variety of response. Some answers might echo aspects such as the "ability to take charge of one's own learning" (Holec 1981). Holec was among the first to define and specify what autonomy was. Parameters of Holec's view of autonomy are the learners':

- engagement and personal motivation to learn;
- ability to determine their own learning objectives;
- ability to define content and progression;
- ability to select from useful strategies or techniques;
- ability to monitor, assess and evaluate their own progress.

These parameters must be related to the specific context of Holec's learners. His intrinsically motivated adult learners of French explored the boundaries of their autonomy willingly and successfully at his CRAPEL institute in Nancy, France. Secondary school learners tend to be less motivated, knowledgeable and experienced than Holec's learners were at the beginning of their studies. Nevertheless, Holec does suggest challenging ways in which autonomy can be fostered in learners effectively, that is by having learners express what they would like to achieve, involving them in ways in which their individual goals can be realised, and having them monitor, assess and evaluate any learning outcome.

Some twelve years later, Kenny (1993) warned against an approach of autonomy that was primarily instrumental. He highlighted the contribution of learner autonomy approaches to human development, saying that "Autonomy is not just a matter of permitting choice in learning situations, or making pupils responsible for the activities they undertake, but of allowing and encouraging learners, through processes deliberately set up for that purpose, to begin to express who they are, what they think, and what they would like to do, in terms of work they initiate and define for themselves." (Kenny, 1993: 440). Definitions of this kind focus on the role of education in human development, where self-expression of feelings, thoughts and identity is seen as a starting point as well as a lifelong goal. Here, moral and intellectual growth are implicitly linked to communication as "a continuous process of *expression*, *interpretation*, and *negotiation* of meaning" (Savignon, 1997:14). Thus, the parameters of Kenny's conception of autonomy are:

- a learning environment set up by the teacher that encourages and enables learners to express and communicate their feelings and thoughts;
- actions and activities that learners initiate and define for themselves, and which are related to what they feel, what they think, and who they are.

Taking Holec's and Kenny's conceptions of autonomy into consideration, we can conclude that they both refer to potentially important parameters of autonomy. Yet, learning environments in secondary education are often far removed from, and may even be hostile to, the situational contexts envisaged by Holec, Kenny and others. Initial definitions of autonomy, especially when they were studied superficially, easily led to a number of misconceptions about what autonomy was. Echoing Widdowson (1990), Little (1991) warns against five "misconceptions" that were current in views of learner autonomy in the 90s. They were the assumptions that (1) autonomous learners make the teacher redundant; (2) any intervention on the part of the teacher may destroy whatever autonomy the learners have managed to attain; (3) autonomy is a new methodology; (4) autonomy represents easily described behaviour; and, finally, (5) autonomy is a steady state achieved by certain learners. The misunderstandings first show how easily the role of the teacher in the learning process can be played down in conceptions of learner autonomy. Merely referring to teachers as *coaches*, *tutors* or *supervisors* does not do justice to the complexities educators face when they are expected to encourage as well as to enable learners how to learn in ways that relate to their personal objectives, abilities and capabilities. At the same time, it is hard to describe autonomous behaviour in objective ways. Autonomy is a complex construct that can only be discussed in relative terms. The misconceptions mentioned by Little are a welcome addition to the parameters we have mentioned so far. It leads to the following demarcations when learner autonomy is fostered in learners:

- teachers are actively involved in creating and monitoring a learning environment that challenges learners to participate;
- teacher interventions may be required and do not necessarily destroy the autonomy the learners have achieved;
- autonomy is not a new methodology, and as such does not replace existing methodologies or techniques;
- autonomous behaviour can neither be described easily nor objectively;
- autonomy is not a steady state, but a dynamic goal to be discussed in relative terms, that is in interaction with other factors.

Little (1997, 1999, 2002) considerably added to an understanding of autonomy as a pedagogical construct. To further specify learner autonomy for educational purposes, Little (1997) defined learner autonomy in relation to particular tasks, highlighting that learner autonomy also consists in the ability to use particular knowledge in and apply certain skills to new tasks situated in contexts different from the learning context and under new conditions. Such a focus on transfer requires learners to (re)consider their task approaches, the strategies to use, or the products to aim for. In the light of these specifications, educators are expected to focus on effective learning activities the learners are to pursue, and on the knowledge and skills that learners need to be successful in relation to a particular task. In this definition, an effective learner is able to transfer the knowledge and skills needed for successful task performance to other tasks or activities, which may well be more complex. When learners have become flexible and adaptive learners in this sense, they can be considered autonomous learners "… when they are able to perform that task (i) without assistance, (ii) beyond the immediate context in which they acquired the knowledge and skills and on which successful task performance depends, and

(iii) flexibly, taking account of the special requirements of particular circumstances." (Little, 1997: 14). Little's specifications of autonomy in pedagogical contexts lead to additional parameters. Learner autonomy:

- is action-based, i.e. seen in relation to particular tasks, actions and activities;
- is transfer-oriented, i.e. consists in the ability to use particular knowledge in and apply certain skills to new tasks in different situations and contexts;
- puts great demands on the knowledge and skills needed to carry out a task successfully as well as on the type of tasks, actions and activities to be carried out.

In 1999, Little specified essentials of learner autonomy of learner autonomy by including explicit reflective questions that learners should attempt to answer: "The basis of learner autonomy is acceptance of responsibility for one's own learning, that the development of learner autonomy depends on the exercise of that responsibility in a never-ending effort to understand what one is learning, why one is learning, how one is learning and with what degree of success; and the effect of learner autonomy is to remove barriers that so easily erect themselves between formal learning and the wider environment in which the learner lives." (Little, 1999:14). Reflective open-ended questions, such as the *what, why, how, when, where, how long* of learning activities may help learners to more effectively realise lifelong learning. Such an interpretation of learner autonomy requires:

- the learners' acceptance of responsibility for their own learning;
- never-ending effort to understand and explicate what one is learning, why one is learning, how one is learning and with what degree of success;

In a later definition, Little, Ridley and Ushioda (2002) underline that the second or foreign language learner's potential to act independently can only be developed if:

- *learners take their first steps towards autonomy when they begin to accept responsibility for their own learning;*
- *they exercise and develop their autonomy by sharing in the decisions and initiatives that give shape and direction to the language learning process;*
- *by planning, monitoring and evaluating their learning, they develop their metacognitive and metalinguistic capacities (their ability to reflect on the learning process, the forms of the target language, and the uses to which the target language can be put).* (Little, Ridley & Ushioda, 2002:31)

As a first condition, the learners' motivation or readiness to accept responsibility for their own learning is once more referred to. Secondly, the authors feel that learners should be involved in decisions about what has to be learned and how learning is best organized. Action, however, is not necessarily initiated exclusively by the learners themselves. Teachers or more knowledgeable or skilful peers may considerably help a learner to widen horizons. Finally, "metacognitive and metalinguistic capacities" are expected to be developed when a learner reflects on the learning process, the forms of language, and the uses in which particular language forms feature. Therefore, two more parameters can be added to the ones presented so far. Learner autonomy in language learners is likely to increase if:

- learners socially share in the decisions and initiatives that give shape and direction to the language learning process;
- their metacognitive and metalinguistic capacities are developed, i.e. their ability to reflect on the learning process, the forms of the target language, and the uses to which the target language can be put.

A focus on learner autonomy assumes teacher and learner roles that may fundamentally differ from what teachers and learners have been used to. Both Kenny (1993), and Little (1991, 1997, 1999) stress the active role of the teacher when autonomy is to be fostered in learners. As a result, a distinction between teacher and learner roles is considered useful. In view of the limited knowledge and skills of the learners, it seems justified to see the teacher as the prime person responsible for creating a challenging learning environment. After all, teachers are supposed to be more knowledgeable and skilful than their learners in the subject they teach. It is important, though, to stress that learners are invited to share in and add to the learning environment initially created by the teacher from the beginning onwards.

We derive thirteen parameters from the interpretations and definitions of autonomy we have so far discussed. Our argument is that if teachers concentrate on one or more of the parameters below and involve their learners in any choices that can be made, LA is likely to increase. The parameters we identify are the:

- ✓ *engagement and motivation* to foster LA of both the teacher and the learner**;**
- ✓ creation and maintenance of a *challenging learning environment;*
- ✓ specific *knowledge, abilities and willingness* required to learn effectively***;***
- ✓ *learning objectives of the curriculum;*
- ✓ *curricular content and pace;*
- ✓ *tasks, actions and activities;*
- ✓ *learning strategies and techniques* involved*;*
- ✓ level of *transference of learning tasks, actions and activities;*
- ✓ *acceptance of responsibility* by the learner*,*
- ✓ learner's *concentration, perseverance and effort;*
- ✓ opportunities for the learner for *self-expression by communication of feelings and thoughts;*
- ✓ ways in which progress is *monitored, assessed and evaluated;*
- ✓ roles of *metacognitive* and *metalinguistic reflection.*

We will now deal with the second question this chapter promised to address.

3.3 Why learner autonomy?

In four sections, we will argue why LA is an acceptable pedagogical goal in formal second or foreign language instruction. We arrive at justifications by looking at autonomy, self-determination and human agency from four perspectives: *socio-historical*, *philosophical, political,* and *psychological.*

3.3.1 A socio-historical perspective

Adopting a sociological as well as an historical perspective, may help us find arguments why LA can be considered a feasible pedagogical goal. From the individual cases of Socrates and Comenius, we will move on to four culture orientations that have been alternately dominant in Western societies. Finally, we will

arrive at a discussion of our present-day knowledge societies. Our socio-historical exploration will result in three justifications of LA as a pedagogical construct.

Socrates and Comenius

Conceptions of autonomy did not grow out of a social and historical vacuum. One of the reasons why autonomy is an acceptable and feasible goal in education is that it has long been central to cultures and their representatives favouring liberal philosophies of education. Among the many historic advocates of learner autonomy, we would like to single out *Socrates* and *Comenius*. Not surprisingly their names have been attached to projects subsidized by the Council of Europe to further education across cultural and political barriers. *Socrates* (469-399 BC), the Athenian philosopher immortalised in the dialogues of his disciple Plato, asserted that true knowledge emerges through dialogue, systemic questioning and an abandoning of any uncritical claims to knowledge. By his engaging Socratic method, he had his learners explore and expand their knowledge and skills on their own by engaging in challenging dialogue. *Comenius* (1592-1670), the Moravian pastor and educationist who believed that universal Christian brotherhood could be achieved through the improvement of education, taught Latin communicatively and involved learners in their learning process by building on their experiences, without exerting his authority as a teacher. He was a productive philosopher and writer. Among his publications were *The School of Infancy* (1630), a handbook for parents and educators, which was the first systematic pedagogical work dealing with pre-school education and *The Gate of Languages Unlocked* (1632), a textbook for teaching Latin in a playful and communicative way. His most famous work is perhaps *The Labyrinth of the World and the Paradise of the Heart* (1631), in which he presented his ideas how society might be improved by the promotion of regular dialogue among nations. Over three centuries later, some of his ideals became reality with the establishment of the United Nations and a united Europe (Savignon, 1997).

Unfortunately, the socio-political environments in which both Socrates and Comenius lived were hostile to their enlightened ideals and ideas. Socrates was forced to commit suicide by the Athenian authorities. Socrates was considered disruptive, because he opposed tyranny and empowered his learners to expose false reasoning. Comenius had to flee his country after the massive re-catholization ordered from Vienna after 1620. Political and religious forces in their communities saw their views of autonomy, learning and education as serious threats to existing power structures.

Fortunately, most educators today are less likely to suffer misfortunes such as suicide or exile when their views of education differ from what is considered as good practice by politicians, inspectors, managers or society at large.

What does this succinct information about Socrates and Comenius tell us about learner autonomy? First of all, that conceptions of learner autonomy are often determined by the socio-cultural context that surrounds the conceiver. At the same time the liberal philosophies favoured by Socrates and Comenius were highly personal and deeply embedded in what they believed to be effective learning and effective teaching. Secondly, the liberal views of Socrates and Comenius sho that learner autonomy is not a new approach to learning. Rather it is an aim that can be attained by a challenging variety of means. Both Socrates and Comenius were ultimately aiming at learner autonomy and social responsibility in their own idiosyncratic ways. Finally, the socially-embedded yet personal views of Socrates and Comenius are an invitation to delve more deeply into the construct of learner autonomy and study it from dialogical, interactive and social perspectives.

What do the individual cases of Socrates and Comenius add to the parameters of autonomy presented above ? They show the importance of the role of the teacher and the importance of quality time between teacher and learner. Socrates and

Comenius shared conception of autonomy as a pedagogical construct and wished to involve their learners in learning. Socrates attempted to foster autonomy by involving the learner, and himself, in a constant search for true knowledge through dialogue and systemic questioning. Comenius attempted to involve learners by starting with their experiences and the thoughts and feelings these experiences had aroused. Thus, individual conceptions of learner autonomy in the past may add to a better understanding of today's thinking about LA. We will continue our socio-historical focus by moving from the individual to the social and discuss the concept of self-determination in four prevalent culture orientations in western societies.

Self-determination in culture orientations

Every society has its own views and ideas about how it should function. In sociological terms, these views and ideas are often referred to as culture orientations. A society may have its own specific culture orientation at a certain point in time, which influences how people think and act in different ways and provides guidelines for their daily actions. As a rule, various culture orientations co-exist within a society, but one orientation is typically the most influential for a given period of time. Matthijsen (1972) distinguishes four culture orientations that have influenced periods within the past millennia:

- The *aristocratic culture orientation.* Its central notion is that the leadership of a state can be entrusted only to an elite of high-ranking people (nobility or patricians).
- The *theocratic culture orientation.*Characteristic of a society that considers God the ultimate authority and is primarily oriented towards priests ruling as direct servants of God.
- The *meritocratic culture orientation*. People derive social status from their individual abilities and achievements.
- The *democratic culture orientation*. The direct or indirect government of a state by the people is at the centre. The individuals in society decide who rules. The democratic orientation explicitly incorporates a model of self-determination. An important aspect of this autonomy model is that groups and individuals are meant to be given opportunities to develop their talents or abilities.

The common factor in the first three orientations is that power in society is beyond the reach of ordinary persons. Power is in the hands of select groups of privileged individuals, such as nobility and clergy. In the fourth orientation, the individual is meant to have social and political status irrespective of ability and achievement. Moreover, interest in the individual is not limited to newly gained social and political status. Societies are concerned with creating opportunities for their members to develop their potential. This aspect of autonomy is most clearly visible in a democratic culture orientation. In Europe, as well as in the Americas, the concept of creating opportunities for self-determination has gradually gained acceptance and, in turn, influenced education, as history has shown.

The nature of education was often passionately debated in the light of changing social perspectives, resulting in movements concerned with all forms of education on both national and international levels. In the Netherlands during the seventies, great interest arose in concepts such as learner autonomy within various forms of socially committed project education. Among other things, this interest was a reaction to classic behaviourism. Rejection of behaviourist determinism translated into a general opposition to the establishment and a search for alternatives. An important objection against the established system of education was its one-sided emphasis on cognitive development. For critics, the system of education was too rigid, both because of the strict separation of primary and secondary education and the system of ability

streaming and tracking from the first year of secondary education onwards. Moreover, links between school learning and the life experiences of the child were generally missing. Among other criticisms levelled were the ineffectiveness of formal education in abolishing social inequality of opportunities and, ultimately, a failure to stimulate and fulfill the needs of learners for independence, responsibility, and participation. (For a summary of these criticisms, see Freire, 1972; Illich, 1979; and Rogers, 1983).

In the early 1980s, people lost interest in socially oriented project education, in which explicit attention was paid to the social relevance of educational content. Once again, education was geared towards professional skills and practical thinking. Interestingly, in the 1990s a two-track educational policy emerged in the Netherlands. One track was related to meritocratic utility thinking and was stimulated by business and industry with the important cemtral concepts of selection and restriction of choices. The other track was related to the renewed interest in an autonomy model to empower every individual in society to develop as a human being responsible for social prosperity and welfare. Certain notions of learner autonomy as guiding principles began to attract the attention of policy makers. The elements for a top-down educational reform were in place. In the Netherlands, it resulted in an educational reform geared at fostering a learner autonomy that has both meritocratic and democratic traits. In chapter 6 we will return to this reform, when we discuss the context of Dutch secondary education.

We have argued that the need for LA stems from what is in essence a democratically oriented view of education. Sometimes a more democratic orientation develops as a reaction against a dominant orientation, as was the case with the aristocratic, theocratic, and meritocratic orientations identified above. The meritocratic and democratic culture orientations were among the more influential in the past two millennia. Only the democratic culture orientation incorporates a model of self-determination, and it is this orientation that has gained influence in Europe in the past few decades. We feel the democratic culture orientation offers a second justification of learner autonomy as a pedagogical goal.

A final socio-historical argument for embracing LA in formal education is offered by the present demands for knowledge and lifelong learning in societies.

Lifelong learning in knowledge-based societies

In knowledge-based societies, it is now generally accepted there is a need for *lifelong learning* that transcends the boundaries of formal education, as well as the limitations of discrete sets of knowledge and skills. Knowledge in present-day societies tends to outdate fast. Many societies are greatly enhanced by the World Wide Web, the Internet and increased mobility, which all have literally opened up worlds of information and knowledge we were unable to imagine until recently.

The basis for designating advanced modern societies as knowledge societies is the transformation of existing societal structures by knowledge. Politicians and economists see the enormous increase of information and knowledge first and foremost as a core resource for economic growth and employment. Educationists and educators increasingly realise that the worlds of knowledge that can be accessed by the learners is both a challenge to the curriculum and an opportunity to develop the autonomy of learners. Costa and Liebmann (1995) had earlier commented on the resulting challenge to education, pleading for an integrative and multidisciplinary approach to knowledge, science and scholarship. They put forward:

> *… that with knowledge doubling every five years – every 73 days by the year 2020 – we can no longer attempt to anticipate future information requirements. If students are to keep pace with the rapid increase of knowledge, we cannot continue to organise curriculum in discrete compartments, ….the disciplines as we have known them, no longer exist.*

> *They are being replaced by human inquiry that draws upon generalised transdisciplinary bodies of knowledge and relationships.* (Costa & Liebmann 1995:23)

In view of the rapidly changing needs of knowledge-based communities, it seems fair to acknowledge that teachers cannot teach everything that learners need to know or would like to learn. This concern is relevant to the foreign language teacher. Foreign languages are used to access information that is only a few mouse clicks away. Foreign language teachers have to ensure that the learners first of all have open and timely access to information and knowledge in the foreign language being taught. In their instructional practices, they should have their learners develop the capacities to access, analyse, interpret, and evaluate relevant information. In addition, teachers should explore the opportunities for the learners to use their newly acquired knowledge to develop as human beings in interaction with others across the frontiers of old.

Concerning the focus discussed above, learners are perhaps better served when they are equipped and supported to actively direct and regulate their own learning. We feel this is a third socio-historical argument that justifies a focus on learner autonomy in formal education.

However, claims for LA seem also supported by insights from other perspectives than our socio-historical stance. In the next section, we will explore autonomy from a philosophical point of view and focus on four mainstream accounts of human agency.

3.3.2 A philosophical perspective

Philosophical studies of human agency and personal autonomy have resulted in four main views or accounts, which partly overlap. Two of these accounts help to justify an LA approach in education.

Philosophical accounts of human agency

Philosophers have long been interested in human agency and personal autonomy. An agent is a person who acts. Action must be initiated. One cannot initiate action without exercising one's power to do so. Human beings have authority over themselves for the simple fact that they can initiate their own actions. From this view, personal autonomy is seen as a mere form of self-government. The authority we have over our actions is but a formal feature of agency. This does not imply that whenever we act, the forces that move us are manifestations of our own power as agents. Our motives are not necessarily related to the decision-making power that moves us to act. Agents determine how to act, but the job can be done without completely being in control.

Philosophers have concentrated on the relationship between the agent's power and the power of the forces that move him/her by attempting to answer two essential questions. What motives are attributable to the agents themselves and what motives can be seen as external? What distinguishes motives on which the agent has conferred authority from the motives whose power has reduced authorisation to a mere formality? Philosophers have proposed different accounts of a human being's special relation to personal motives. Buss (2002) mentions four more or less overlapping approaches to personal autonomy: *coherentist*, *reasons-responsive*, *responsiveness-to-reasoning* and *incompatibilist*. The four approaches may help us unravel some of the complexities between a person's motives and the power to act. The relationship between a person’s motivation and motives to act and his/her power to act is at the heart of our exploration of autonomy.

Coherentist approaches

In a coherentist approach, it is argued that human beings can only be called autonomous agents if they accept their motives, identify with them, approve of them, or believe they make sense in relation to any long term commitments or plans they may have. The approach assumes a reflexive attitude towards the motives that make a person act. If a person endorses and accepts these motives, that person is called an autonomous agent. Actions occur with the permission of the agent, if not necessarily at his or her command. If people do not accept or identify with the motives that make them act, then these motives are caused by external forces in conflict with their personal autonomy.

Reasons-responsive approaches

According to philosophers advocating a reasons-responsive conception of autonomy, human beings are not seen as autonomous agents unless their motives, or the mental states that produced them, are responsive to a wide range of reasons for and/or against behaving as they do. It is considered that human beings who fail to appreciate and evaluate a variety of reasons for action are unlikely to govern themselves well. Ignorance of "... a pattern of actual and hypothetical recognition of reasons (some of which are moral reasons) that is understandable to some appropriate external observer" is seen as a serious threat to self-government (Fischer and Ravizza, 1993:90).

Responsiveness-to-reasoning approaches

According to responsiveness-to-reasoning theorists, people govern themselves if they have the capacity to evaluate their motives by relating them to whatever else they believe in or desire, and adjust these motives in response to their evaluations (Buss, 2002:4). Responsive-to-reasoning accounts, with their focus on the reasoning process, often suggest that self-government requires the capacity for self-transformation and development. On this assumption, autonomous agents are capable of changing their minds when they discover good reason to do so in the course of their reasoning. It is important to note that changes of mind caused by sensitivity to reasoning do not necessarily lead to a growth in autonomy of the agent in question.

Whether a person reasons correctly or wrongly is norm-referenced and determined by the reasons or criteria that underlie a particular norm. The norms that govern the reasoning process seem to be shaped both by internal and external forces. In a responsiveness-to-reasoning approach, a person who is reasoning is likely to investigate the internal and/or external forces that are involved and how these forces interact in the course of the reasoning process.

Incompatibilist approaches

A fourth view of personal autonomy is generally referred to as incompatibilist. According to this account it is impossible to validly distinguish between internal and external forces that influence a human agent's intention-forming process. Incompatibilists feel that autonomy is incompatible with determinism. One of the more rigorous expressions of the incompatibilist position is to be found in Kane (1996). According to Kane, the wish to be an autonomous agent is the desire to have "the power to be the ultimate producer of [one's] own ends ... the power to make choices which can only and finally be explained in terms of [one's] own [will]". "No one," he argues "can have this power in a determined world." (Kane, 1996: 254).

The four mainstream philosophical accounts of autonomy first of all invite us to critically explore human agency and conceptions of personal autonomy. Second, two of the four accounts implicitly embody views of autonomy that are relevant to and actually justify an exploration of autonomy in pedagogical contexts. The *reasons-responsive approaches* stress the importance of individual choice from a wide variety

of motives. The assumption here is that the more varied a person's motives are, the more autonomous that person is in his/her activities and actions. Teachers can help learners explore and develop such a wide array of motives. The *responsiveness-to-reasoning* approaches focus on the quality of reasoning before action is taken. Language teachers can trigger a learner's motives as well as the quality of reasoning. This can e.g. be done by asking question such as: 'What is the role of vocabulary and idiomatic expressions when you learn to communicate in a foreign language?'; 'Why is it that in order to really understand a written or oral text, you need to know at least 8 out of any 10 words of it?'; 'How are you going to make sure you will master at least 80% of the words and phrases of a given written or oral text?'; 'What can you do to retain vocabulary and idioms that you once studied?'.

Thus, adopting a philosophical perspective on autonomy is a fourth reason that justifies an exploration of autonomy as a pedagogical construct. Such a perspective may lead to viable and practical ways in which learner autonomy can be increased.

Additional justifications for an LA focus can be found if we consider the construct from a political point of view.

3.3.3 A political perspective

Benson (1997) opts for an approach that is politically oriented. He discusses three basic versions of autonomy serving as 'ideal' constructs. The first is a technical version, in which the act of learning a language takes place outside the framework of an educational institution and without the intervention of a teacher. In this version the most important issue of learner autonomy is how to equip the learners with the skills and techniques they need to cope with such situations. A second version of autonomy is psychological. It considers learner autonomy as a capacity- a construct of attitudes and abilities - that allows learners to take more responsibility for their own learning and in which autonomy is developed by internal transformation within the individual learner. The third version is political. At issue is how far a learner has control over the processes and contents of learning. A political version of autonomy highlights the importance of creating structural conditions that enable learners to not only control their own learning but also partly determine the conditions of learning.

Benson reviews three loosely-defined theories of learning, i.e. *positivism*, in which knowledge is considered as the more or less accurate reflection of objective reality, *constructivism*, in which knowledge is seen as the co-construction of meaning, which helps each learner to construct extended versions of reality, and finally *critical theory*, in which knowledge is constructed rather than acquired within the social contexts and constraints of often conflicting ideologies. Benson elaborates on the political version of learner autonomy and on critical theory. In critical theory, learning in general and the learning of foreign languages in particular are dependent on and determined by political and social conditions (see also Pennycook, 1997, 2001). Therefore, Benson (1997) proposes to concentrate on the social and political context in which second or foreign language learning takes place, and on the roles and relationships of the people who interact both inside and outside of the classroom. Benson feels the learner must adopt a critical attitude towards the learning goals, content of learning (e.g. materials and learning tasks) and didactic procedures, in short, towards all areas in which autonomy can be promoted. Benson's political version of autonomy (1997) has origins similar to the critical pedagogy of e.g. Giroux (1981), Mclaren (1989) and Kanpol (1994, 1997). They also criticise formal education, acknowledge the political aspects of schooling and curricula, and attempt to encourage students to develop their own democratic alternatives.

Critical pedagogy is not without opponents. Pennycook (2001) has summarised critiques of critical pedagogy. It has been accused of remaining at the level of 'grand theorizing', that is of remaining unrelated to pedagogical practice. It tends to be prescriptive in the sense of telling teachers what to do, often without being explicit. And, according to the critics, critical pedagogy has been concerned more with North-

American individualistic idealism than with effecting social change (Pennycook, 2001: 130-132). Pennycook acknowledges these criticisms, but, at the same time, recognizes the merits of critical pedagogy, such as offering insight into and understanding of educational theory and adopting a constructively critical approach to language and education. Therefore, he argues for what he calls a "postcritical pedagogy" of language education, which "attempts to deal with the postcolonial challenge of dealing with the Other, the poststructuralist requirement to understand how discourses operate across multiple sites, constructing our world and subjectivities, and the postmodern challenge to deal with the particularities and complexities presented by trying to take differences seriously" (Pennycook, 2001: 140).

The postcritical approach to foreign language learning and teaching advocated by Pennycook is strongly related to the so-called "postmethod" pedagogy of Kumaravadivelu (2001). It was proposed in response to the alleged limitations of the concept of *method* and the transmission model of teacher education. Postmethod pedagogy is based on three principles. The first is *particularity*, meaning that language pedagogy must be sensitive to a particular group of teachers teaching a particular group of learners pursuing a particular set of goals within a particular institutional model embedded in a particular sociocultural milieu". The second is *practicality* that emphasizes the need for teachers to theorize from their practice and practise what they theorize. The third principle is *possibility*, referring to creating opportunities for foreign language learning and teaching "by acknowledging and highlighting students' and teachers' subject positions - that is, their class, race, gender, and ethnicity" and stimulating them to question the status quo that keeps them subjugated" (Kumaravadivelu, 2001: 542).

To conclude this section, a political perspective on autonomy as well as understandings from postcritical and postmethod pedagogy help to critically review pedagogical approaches that aim to foster learner autonomy in learners. They particularly help to highlight, discuss, and negotiate power dimensions, such as certain situational demands or pressures that help or hinder learner independence. As such, a political perspective is a fifth justification for an autonomy focus in formal education.

However, a version of learner autonomy that is exclusively political may not do justice to the socio-historical and philosophical perspectives discussed above. In addition, as was acknowledged by Benson, learner autonomy is a pedagogical construct that can also be considered from a psychological point of view. It is perhaps this psychological angle that provides the most powerful justification why autonomy should be aimed for in formal secondary education.

3.3.4 A psychological perspective

We derive a sixth and final argument for an LA focus from insights gained by psychologists into the human mind and into learning processes. Understandings obtained from humanistic psychology, cognitivism and constructivism together provide an informed view of how learners learn and of how they may learn more effectively. Taking account of these theoretical frameworks may incite teachers to explore the boundaries between what learners can do on their own, and for what actions or tasks they need assistance and support. We will successively refer to humanistic psychology, cognitivism and constructivism and relate these broad theories of mind to what they have all been aiming at: motivating and empowering learners to become more autonomous. Therefore, the relationship between motivation and autonomy will be explored in a section that closes off our final argument.

Humanistic psychology

Humanistic psychology opposed behaviourism, which had dominated educational psychology for decades. Classic behaviourism reflected a strongly deterministic and mechanistic portrayal of man. External regulation of learning behaviour left the learner little opportunity to experience independence and responsibility in his/her individual learning process (Watson 1913, Thorndike 1932; Hull 1943). Thus, classic behaviourist learning perspectives did not foster learner autonomy. In the 1960s a number of theories of human learning, notably humanistic, cognitive and constructive views, were developed in reaction to behaviourist perspectives. More recent versions of behavioural theory on learning still resonate in discussions of self-regulated learning, such as Skinner's views of operant conditioning (1957, 1971), with its focus on positive reinforcement of desired behaviour, or Bandura's social learning theory, which spans both behavioural and cognitive frameworks (1971, 1986).

Humanist psychologists acknowledge that stimulus-response reactions may play a role in learning processes. However, they feel that behaviour and experience are primarily initiated by the individual and not exclusively caused by external incentives. Humans have the principal and unique capability to make choices and to distinguish themselves from one another. They assume responsibility for their choices. Humanists thus emphasize the ways in which learners perceive their environments. In education, humanist psychologists attribute an important role to a learner's thoughts, feelings, and motivation. Two representatives of humanist psychology who have been explicit on educational matters and learner motivation are Maslow (1970) and Rogers (1983).

Maslow (1970) saw motivation as a construct in which the ultimate attainment of goals was only possible by passing through a hierarchy of needs. He saw motivation as first of all dependent on the satisfaction of fundamental physical necessities, such as air, water and food. Having satisfied these basic needs, the learner's needs are more social. The fulfilment of senses of community, security, identity, and self-esteem finally lead to what Maslow calls *self-actualization*. We will consider Maslow's theory in more detail when we discuss theories of motivation.

Rogers (1983) saw humans primarily as affective and emotional beings, and not only as physical and cognitive persons. His formal principles focused on the internal forces that cause a person to act: the individual's self-concept and his/her personal sense of reality. Rogers felt a non-threatening environment was important for a person to form a picture of reality congruent with the way things are. Only then will human beings grow and learn, and develop into "fully functioning persons". Rogers developed the theory of experiential learning. He distinguished cognitive learning from experiential learning, viewing the latter as more meaningful and significant. Cognitive learning relates to academic knowledge such as learning vocabulary or multiplication tables and the latter corresponds to applied knowledge, such as learning about writing strategies to become a better writer. Experiential learning explicitly addresses the needs and wants of the learner. Rogers lists as qualities of experiential learning: personal involvement, self-initiation, self-evaluation and pervasive effects on the learner. To Rogers experiential learning is equivalent to change and growth. He feels that all humans have a natural propensity to learn. It is the teacher's role to facilitate such learning, which includes (1) setting a positive climate for learning, (2) clarifying the learner's goals (3) organizing and making available resources, (4) balancing intellectual and emotional components of learning, and (5) sharing feelings and thoughts with learners without teacher domination. According to Rogers, learning is facilitated when: (1) the student participates completely in the learning process and has control over its nature and direction, (2) is primarily based on direct confrontation with practical, social, personal or research problems, and (3) self-evaluation is the principal method of assessing progress and success. Rogers also emphasizes the importance of learning how to learn and the

development of openness to change. Humanists such as Rogers build on the assumption that if a teacher manages to engage the learners and succeeds in boosting and building their confidence, there will be a firm base for the learners to explore and extend the boundaries of their autonomy.

Maslow's concept of self-actualization and Rogers's view of the "fully functioning person" have important implications for education. The focus is away from teaching to learning. Learning how to learn is considered more important than being taught something from the superior vantage point of a teacher who decides what must be taught and learned. In order to have learners learn, teachers establish interpersonal relationships with their learners characterized by open-mindedness, sincerity and the teacher trusting, accepting and prizing the pupil or student in open and empathetic communication. Such implications help to modify the parameter of learner autonomy under teacher roles, which we have called *challenging learning environment*.

Unfortunately, transfer of Maslow's and Rogers's ideas to educational practice may easily lead to misunderstandings. Teachers may take their non-directive approaches too far, to the point that too much time is spent on allowing the learners to discover facts and principles for themselves. In addition, teachers may take the focus on the non-threatening environment to the point where the necessary facilitative tension and challenge to learn is absent. There is ample research documenting positive effects of a type of classroom competitiveness that stimulates instead of damages self-esteem and the motivation to learn (Bailey, 1983). A potential remedy to these misunderstandings may be the adoption of a more rational approach to the human mind and to learning, such as cognitivists have done.

Cognitivism

Cognitive psychologists (Bartlett 1932; Anderson & Ausubel 1965; Neisser 1967; Ausubel 1968, Rumelhart & Ortony 1977; Schank & Abelson 1979; Anderson & Bower 1980) approach the human mind from a more rationalist angle than humanists do. They view learners as rational human beings who can use a network of relations, so-called schemata, between different interdependent concepts to understand, store and (re)construct information. Cognitivists are primarily concerned with the mental processes that underlie human behaviour, such as thinking, reasoning, decision-making, and to some extent motivation and emotion. Cognitivism examines questions about the workings of memory, attention, perception, knowledge representation, reasoning, creativity and problem-solving.

Cognitivists generally feel that learners are likely to learn if they are put in a rich and meaningful learning environment that allows them to relate new information to already existing knowledge, skills, concepts or propositions in systematic ways. As an example of a cognitivist theory that has important implications for second and foreign language learning. we will discuss Ausubel's meaningful learning theory.

Central to Ausubel's theory is that learning is seen as a conscious and meaningful process, in which the learner relates new knowledge, skills or understandings to what (s)he already knows, is able to do, or understands. Meaning is not a random, haphazard or implicit response, but a "clearly articulated and precisely differentiated conscious experience that emerges when meaningful signs, symbols, concepts, or propositions are related to and incorporated within a given individual's cognitive structure on a nonarbitrary and substantive basis." (Anderson & Ausubel, 1965:8). This very relatability, according to Ausubel, accounts for at least three phenomena when learners learn: the acquisition of new meanings (knowledge, skills or understandings), the way in which the knowledge, skills or understandings are hierarchically organized in the brain (subsumption into existing cognitive structure), and the level of retention of what is hierarchically stored in the brain (systematic forgetting and language loss or attrition). We will discuss these three phenomena and, while doing so, attempt to show that cognitivism has important

implications for the ways in which second or foreign languages are taught and learned.

According to Ausubel, new knowledge, skills or understandings are acquired if a learning situation is meaningful. He feels any learning situation can be meaningful if (a) learners have a disposition to relate the new learning task to what they already know, and (b) the learning task itself is potentially relatable, and therefore meaningful, to what the learners already know or are able to do. Both the teacher and the learner share the responsibility for the learner's meaningful mind sets and meaningful learning tasks to be carried out. The teacher, however, is expected to prime the learner to meaningfulness if a learner fails to see the meaningfulness of a task. Meaningfulness can be "manufactured" (Smith, 1975:162), that is teachers and learners can make tasks meaningful if they feel the need to and if they are strongly motivated to do so.

Meaningful learning can best be explained in comparison with rote learning, or the learning by heart of discrete elements that do not bear any relationships to existing cognitive structures in the brain. An example would be the rote learning of telephone numbers if the numbers are seen as arbitrary entities that do nor bear any relationship with existing cognitive structures. In other words, the numbers are not *subsumed* under a more inclusive conceptual system. Without repeated conditioning, the numbers will no longer be remembered. The situation is likely to be different if we are asked to memorize addresses. This is because addresses are often associated with physical images of the places in question, the directions that have to be taken in order to get there, the people that live there, or the association with memorable activities a person has come to like or dislike.

Ausubel's theory provides a plausible explanation for the phenomenon of forgetting. Self-contained, isolated entities, such as the translation of difficult words in a text that help a learner to understand what the text is about, will only be retained if they bear some relationship with what the learner is already able to recognize or use, or if the words recur in other texts or vocab lists to come. If learners, on the other hand, have learned to relate unknown idioms to parts or concepts they do understand, or to paraphrase these idioms, give examples of what they refer to, or actively use them in different contexts, then the idioms might be retained for a longer period of time. Meaningful subsumption of material from the outset by the learner as well as by the teacher, will considerably enhance the retention process. According to Ausubel, what will happen with meaningfully subsumed knowledge is that it will be "forgotten" systematically. This means that the meaning of a single item is forgotten to the benefit of an overall structure of framework that includes the meaning of the item. This type of " forgetting" is part of the subsumption process, which is characterized by "memorial reduction to the least common denominator" (Ausubel 1963:218). A child may first be confronted with heat and pain if (s)he first touches a hot fireplace or is constantly warned not to touch that fireplace. Gradually, the child experiences more items that generate potentially painful heat, such as hot coffee, a pan of boiling water, a stove, an iron, or a candle. Now the child begins to form the a general concept of "hotness" and tends to forget all of the individual items that had caused pain, to the benefit of the general concept and future items (s)he will meet that generate heat, such as the exhaust pipes of Daddy's motorbike after a long and fast ride. This process is called *systematic forgetting* or *cognitive pruning* (Brown, 1972).

Cognitivism has important implications for educational practice. It first of all helps us understand how knowledge is possibly processed and retained. It also sensitises practitioners to meaningful tasks, actions and activities that help to create an effective learning environment. Finally, under the influence of cognitivism, the nature of teacher instruction gradually changed. Teacher instruction from a cognitive perspective demanded extended models of direct teacher instruction in which the importance of empathic and meaningful teacher-learner interaction and information

processing by the learners were highlighted. Cognitivist ideas were incorporated in prevailing models of instruction used in foreign language pedagogy. These models are generally referred to as *Direct Instruction* or explicit teaching. According to Rosenshine (1986) direct instruction "is a systematic method for presenting material in small steps, pausing to check for student understanding and eliciting active and successful participation from all students." (1986: 60). The demonstrated effectiveness of direct teaching seems to account for its renewed popularity, especially in the field of cognitive performance (Rosenshine & Stevens, 1986; Good & Brophy, 1991).

We will now turn to constructivism, which can be seen as a more radical school of cognitivist thought that explicitly aims at learner autonomy.

Constructivism

Constructivists have approached human learning from yet another perspective, emphasizing that a human being grows up in a social world and that learning occurs in part through interaction with others. This continuous interaction allows human beings to attribute meaning to the world around them. Constructivists view learning as a continuous interactional process in which learners acquire new knowledge in their own subjective ways, process it, and locate it in the existing structures of their knowledge, experiences and beliefs. This way, learning can be seen as an ongoing process of construction and reconstruction (Boekaerts & Simons, 1995; Lowyck & Verloop, 1995).

Constructivism is the common name of a diversity of theories that are rooted in cognitive psychology. It should be noted that constructivism is a very broad framework in philosophy and science, inspired by concepts and ideas of a host of theorists who were responsible of theories in their own right. Among the theorists who have most notably influenced constructivism are Piaget (1896-1980), with his theory of *genetic epistemology*, and Vygotsky (1896-1934), with his *sociocultural theory*.

We will first discuss what constructivist theories have in common. Next we will turn to a brief discussion of the concepts and ideas of Piaget and Vygotsky. Finally, we will highlight characteristics of a constructive classroom focussing on learner autonomy.

According to Mahoney (2004), five basic themes pervade the diversity of theories expressing constructivism. These themes are (1) active agency, (2) order, (3) self, (4) social-symbolic relatedness, and (5) lifespan development.

Constructivists first of all believe in active agency and dismiss deterministic views that cast humans as passive pawns in the play of larger forces. In a second common contention, constructivists argue that human activity often concerns ordering processes. Humans attempt to order experience by means of tacit, emotional meaning-making processes. Third, there is the contention that personal activity is organised in fundamentally self-referent or recursive ways, which adds to a human's sense of personal identity and selfhood. A fourth common theme is that individuals cannot be seen as separate from their environments. The "self is not an island of Cartesian mentation" (2004:3). As human beings we are firmly embedded in the social and symbolic systems that surround us. These first four themes unite in the last, that is lifespan development:

> *Finally, all of this active, meaningful, and socially-embedded self-organization reflects an ongoing developmental flow in which dynamic dialectical tensions are essential. Order and disorder co-exist in lifelong quests for a dynamic balance that is never quite achieved. The existential tone here is unmistakable. Together, then, these five themes convey a constructive view of human experience as one that emphasizes meaningful action by a developing self in complex and unfolding relationships.* (Mahoney, 2004:3)

A distinction between cognitive and social constructivism is often found in the literature on constructivism. Cognitive constructivism essentially relies on Piaget's genetic epistemology, whereas some of Vygotsky's concepts and ideas have influenced social constructivism. It is important to state, though, that Vygotsky's cultural-historical views of the human mind, often referred to as sociocultural theory, do not simply equal social constructivism. Sociocultural theorists feel that Vygotsky's cultural-historical psychology should be studied as a theory of mind with a number of related core concepts of its own. Bruner has attempted to translate his constructivist view of human experience into a theory of instruction. We will discuss some of the main concepts and ideas of Piaget's and Vygotsky's theories.

Piaget's genetic epistemology

Piaget's theory is generally considered an important contribution to cognitive constructivism. With his backgrounds in biology and philosophy, Piaget conducted programmes of naturalistic research on how knowledge developed in human organisms. Cognitive structure is essential to his developmental theory. He saw cognitive structures as patterns of physical or mental action that underlie specific acts of intelligence and correspond to stages of child development. He identified four primary cognitive structures: the *sensorimotor stage* (0-2 years), in which intelligence takes the form of motor actions, the *preoperation period* (3-7 years), which is intuitive in nature, the *operational stage* (8-11 years), which is based on logic but depends on concrete referents, and finally the stage of *formal operations* (12-15 years), in which thinking involves abstractions.

According to Piaget, cognitive development consists of a constant effort to adapt to the environment in terms of assimilation and accommodation. Assimilation involves the interpretation of events in existing cognitive structures, whereas accommodation refers to necessary changes in the cognitive structure in order to make sense of new experiences, events, knowledge or skills. This makes Piaget's theory inspirational to other constructivist perspectives on human learning.

Kearsley (2002) highlights four piagetian principles relevant to the field of education: (1) Children will provide different explanations of reality at different stages of cognitive development. (2) Cognitive development is facilitated by providing activities and situations that engage learners and require adaptation (i.e. assimilation and accommodation). (3) Learning materials and activities should involve the appropriate level of motor or mental operations for a child of a given age; tasks that go beyond current capabilities should be avoided. (4) Methods should be used that actively involve learners and present challenges.

Standard discussions of the differences of cognitive and social constructivism focus on the locus of cognitive development. For Piaget, individuals construct knowledge through their actions on the world. By contrast, the Vygotskian claim is that understanding is essentially social in origin.

Vygotsky's sociocultural theory

Vygotsky's sociocultural theory (SCT) is a theory of mind that emphasises that human beings grow up in a social world and that learning occurs through interaction with a variety of sociocultural phenomena. It acknowledges the central role that social relationships and culturally constructed concrete (e.g. computers) or symbolic (e.g. language) artefacts play in organising uniquely human forms of thinking. The summary of Vygotsky's life and work has been largely based on Valsiner & Van der Veer (2000). Consulted sources for the main concepts and ideas of SCT have been Van Lier (1996), Lantolf (2004), Grabois (2004), Thorne (2004) and Lantolf & Thorne (2006).

It is important to state that Vygotsky's concepts and ideas can only be understood in relation to the Marxist context of dialectical materialism of the Soviet Union in the 1930s. Vygotsky aimed at developing a new psychology of practical relevance to education and the advancement of the Soviet Union. Born of Jewish parents in Orsha, he grew up in Gomel, the second-largest town of the present Republic of Belarus. At Moscow university Vygotsky's breadth of knowledge was shown by the courses he pursued in psychology, medicine, law, philosophy, education, and, above all, in his first love literature. His interests in history and psychology grew from his literary concerns, most notably from the emotional responses people experienced when reading a work of literature or viewing a piece of art (Valsiner & Van der Veer 2000: 330). He became a teacher and teacher educator of primary school teachers. Together with his students, he carried our psychological experiments on which he, at the age of 28, reported at the Second All-Russian Congress of Psychoneurology in Leningrad in 1924. Not long after his presentation, he joined the Institute for Psychology in Moscow as a junior lecturer and started his Ph.D. dissertation *The Psychology of Art.* Eventually, he joined the Institute's subsection for pedology and defectology. He is said to have teamed up with N.S. Krupskaya, Lenin's wife. Together with a number of colleagues they took up a massive reform of the Soviet educational system. The reformers objected to the training of "purely mechanical skills at the expense of broad knowledge" (Prawat, 2000: 693) and instead attempted to empower learners "to lift themselves through education" (ibid.)

Sadly, the new psychology Vygotsky had in mind for the reform of education and the advancement of the USSR fundamentally differed from what Stalin considered right and proper. Stalin had insisted it was impossible for thoughts to be independent of language. This view was to Stalin's advantage, because if he could control "what people said he could control what they thought" (Prawat, 2000: 691). Vygotsky's argument was that "thought and language were separate but dialectically intertwined features of higher forms of consciousness" (Lantolf, 2004: 14). Vygotsky felt the human mind was mediated by symbolic artefacts such as language, just as labour activity was seen to be mediated by tool use in Marx's writings. In 1931, after endless questioning, Vygotsky and his colleagues were severely censured, politically isolated, and their ideas dismissed as non-materialist and bourgeois, based on capitalist thinkers such as Dewey and Piaget. Not until 1934, Vygotsky was allowed a brief restart as head of the Institute for Experimental Psychology. In June 1934 he died of tuberculosis. For decades to come his sociocultural theory was to be ignored. This radically changed when new Russian editions of his work were published, in addition to a growing number of translations in other languages. It led to a revival of sociocultural theory and proved to be an inspiration for what some scholars have called social constructivism.

In order to better understand Vygotsky's cultural-historical psychology, we will briefly discuss his core concepts *mediation*, *internalization*, the development of *social*, to *private* to *inner* speech, *imitation*, and the often quoted *Zone of Proximal development* (ZPD).

Mediation

One of the core concepts of SCT is that the human mind is mediated. Mediation is used in the sense that higher forms of mental activity, such as planning, voluntary attention, intentional memory, rational thinking, or learning, are enabled or affected by symbolic artefacts, such as numbers, graphs, and above all speech and writing. These symbolic artefacts are "culturally constructed and are passed on and appropriated, often in modified form, from one generation to another" (Lantolf, 2004: 15). Just as with physical tools, a symbolic tool such as language establishes "an indirect, or mediated, relationship between ourselves and the world" (Lantolf, 2004: 15). A shovel is an example of a physical tool that helps a person dig a hole more

efficiently. In similar ways graphs, numbers or language may help to plan the kind of hole that needs to be dug before the actual digging is started. In such cases, mental activities precede the actual physical activity of digging. The relationship between mental activity and a concrete object is reciprocal. Words in themselves can be seen as concrete objects when spoken or written. They generally become meaningful when they are expressed, listened to, or read in a specific practical context. Words do not only play a role in planning concrete activities such as digging a hole, but in Vygotsky's theory they are the starting point of concepts and conceptual thinking. Vygotsky felt language to be related to developmental stages. Initially, activity precedes language, which at a later stage reverts to language preceding an activity.

Internalisation

Internalisation is a person's ability to perform a concrete or mental activity by way of mental representations of physical objects, problems or certain situations in the real world. Such a person is no longer dependent on the actual presence of physical objects, problems or situations. Wertsch (1998) offers the example of the difference between a novice and an expert pool player. The novice must actually play a shot in order to determine what the result of a shot is likely to be. The novice solely relies on the external material support of the cue stick, the balls and the table. The expert player is able to determine the outcome of a shot before actually playing it by visualizing the shot in his/her mind. Internalisation is the name for the processes through which interpersonal and person-environment interaction forms or transforms a human's internal mental functions. It is a process through which the specifically human forms of thinking, what Vygotsky refers to as higher forms of consciousness, are created. Just as was the case with mediation, the process of internalisation and their corresponding forms of higher consciousness are culturally determined. Kozulin (1998) offers the example of an Ethiopian immigrant who was asked to group a number of musical instruments. The person could think of no other way than to classify the instruments in groups that are used together on particular social occasions. A classification into e.g. wind instruments, strings, percussion etc. was not in any way part of the person's cultural background. The formal schemata on which objects are classified seem culturally constructed.

From social to private to inner speech

In Vygotsky's view, internalization is a concrete activity that is manifested in a developmental shift from *social dialogue* (talk between two ore more people), to *private speech* (a person talking to him/herself, creating a dialogue between the 'I' and 'Me'), to *inner speech* (speech that is no longer observable in form and content). The transition from social to inner speech is characterized by talk that is increasingly elliptical in form and meaning up to the point where it is no longer observable. Because private speech derives form social speech and is a precursor to inner speech, Vygotsky theorized that studying private speech would help to reveal a person's mental development. Vygotsky and his colleagues carried out studies of the ways in which children developed the ability to use private speech to regulate their own mental and even physical activity as they carried out concrete, culture specific tasks (Vygotsky, 1987, Luria, 1982). In the case of difficult tasks important enough to see through, subvocal inner speech may become vocal again. The internal mental operations of inner speech revert to observable private speech. This may lead to that person seeking help from others or using artefacts such as a manual, calculator, or a computer. Frawley and Lantolf (1985) referred to this process as *re-accessing.*

Imitation

Imitation is fundamental to the process of internalisation (Vygotsky, 1987: 210). Imitation is generally seen as repetition or mimicking, i.e. copying what has been

heard or seen without any changes or additions. Vygotsky, and with him Newman and Holzman (1993) and Tomasello (1999), see imitation as a potentially transformative activity. Imitation is seen as a unique form of cultural transmission that enables the imitator, e.g. a child learning how to use a language, to change or adapt a message or activity. Imitation of this nature may imply agency and intentionality, unlike many phenomena in nature, such as seasonal changes, waves breaking on a beach, or the earth orbiting around the sun (Lantolf, 2004: 23).

Zone of Proximal Development

Without any doubt the most well-known concept of SCT is the zone of proximal development or ZPD. This metaphorical area frames the level of potential development a learner can achieve "through problem solving under adult guidance or in collaboration with more capable peers." (Vygotsky 1978:86). The ZPD is a metaphor for the difference between what a learner can achieve when acting alone and what that person can accomplish with support form someone else and/or cultural artefacts. It was originally proposed by Vygotsky as a way to capture "the process through which institutionalized schooling impacts on intelligence, as measured by IQ tests" (Lantolf, 2004: 24). The ZPD is a concept different from notions such as Wood, Bruner & Ross's *scaffolding* or Krashen's *i + 1* hypothesis, which are based on interlinguistic competence that is yet to be internalised by a learner. Van Lier (1996, 2000) states that for a proper understanding of the ZPD it is important to realize that the metaphor is based on the potential development of the learner, which he visualizes as a small circle of self-regulation (what a learner can do, understand or regulate on his/her own), surrounded by a bigger circle (what a learner can potentially achieve with the help of external sources). The areas between the two circles is the ZPD.

There are two basic interpretations of the ZPD. One interpretation narrows down the ZPD to the necessity of an 'expert' leading the 'novice' for a learner to develop. A broader interpretation of the zone also allows for the fact that peers working jointly at a task can co-construct new knowledge and expertise as a group feature. In a broader interpretation of the ZPD it is conceived as the collaborative construction of opportunities, labelled as *occasions for learning* (Swain & Lapkin, 1998) or as *affordances* (Van Lier 1996, 2000). According to Van Lier (1996: 192-194), interaction is needed for a learner to explore his potential development. This could be interaction with 1) (re)sources of one's own, such as books, computers or calculators; 2) less knowledgeable or skilful peers; 3) peers with similar knowledge or skills; 4) more knowledgeable or skilful peers or experts. The preferred type of interaction is dependent on factors such as learner characteristics, situational demands, language ability, or the learning task on hand. Van Lier refers to these as *multiple zones of proximal development* and relates these zones to the metaphor of *scaffolding*. Wood, Bruner and Ross (1976) used the notion of *scaffolding* to provide a framework with concrete suggestions that enables teachers to create appropriate conditions for the construction of a learner's ZPD. According to Wood et al. the major features of scaffolding are (1) *recruitment*, or focusing the learner's attention to the task, (2) *reduction in degrees of freedom*, or task simplification to limit the cognitive demands, (3) *direction maintenance*, primarily related to learner motivation, (4) *marking of critical features*, (5) *frustration control*, and finally (6) *demonstration.* Scaffolding assumes a teacher role that is more active and complex than the labels of the teacher as a mere 'guide' or 'facilitator' suggest. It would seem that the notion of scaffolding is inextricably bound with the notions of assessment and evaluation. Teachers have to be able to assess what particular knowledge or skills are required in order to empower learners to explore their ZPDs and develop additional autonomy, knowledge and skills in the process.

Vygotsky's sociocultural theory has been inspirational to a host of theories and language pedagogies. *Activity theory* is an example of a theory that has directly

grown out of Vygotsky's SCT. It is a later development within Vygotsky-inspired research that emphasizes cultural activity as the leading principle that relates external forms of social life to individual and collective psychological functioning. For a more detailed discussion of activity theory and its relation to L2 development and research see Leont'ev (1978) and Lantolf & Thorne (2006).

Constructivist classrooms

Both Piaget's and Vygotsky's theories have been inspirational to constructivist theories. Brooks & Brooks (1993) have hallmarked twelve principles that guide the work of a teacher in a constructivist classroom. A constructivist teacher:

- encourages and accepts learner autonomy and initiative;
- uses raw data and primary sources along with manipulative, interactive and physical materials;
- uses cognitive terminology, such as *classify, analyse*, and *create*, when framing tasks;
- allows student responses to drive lessons, shift instructional strategies, or alter content;
- inquires about students' understandings of concepts before sharing his/her own understandings of those concepts;
- encourages students to engage in dialogue both with the teacher and with one another;
- stimulates student inquiry by asking thoughtful, open-ended questions and encouraging students to ask questions of each other;
- seeks elaboration of students' initial responses;
- engages students in experiences that might engender contradictions to their initial hypotheses and then encourage discussion;
- allows waiting time after posing questions;
- provides time for students to construct relationships and create metaphors;
- nurtures students' natural curiosity through frequent use of learning cycle models.

Constructivist classrooms are not without success, as the following research outcomes show**.** McCormick & Donato (2000) explored the usefulness of questions during teacher-fronted activity in an ESL classroom. They studied whether the questions asked might provide appropriate scaffolding to facilitate learning. They found positive evidence in teacher questions in relation to many of Wood et al.'s features of scaffolding, particularly as regards recruitment, reduction in degrees of freedom, and direction maintenance. In a related study, Nassaji & Wells (1998) investigated that traditional IRE (initiation-response-evaluation) or IRF (initiation-response-feedback) dialogic patterns did initially facilitate learning, and helped to encourage a more distributed and socially-constructed learning environment. At the same time research has, often convincingly, shown that second or foreign language learners are also capable of successfully exploring and expanding their own ZPDs in collaboration with others. Donato (1994) looked at peer interactions of French students to determine whether they would successfully construct a scaffold by sharing their understandings and negotiation. Donato found that students working in peer dyads are indeed capable of creating mutually constructed scaffolding created in expert/novice pairs, particularly when it concerned matters of grammaticality. This research has important implications for classroom learning environments, particularly as regards the value of pair and group work (Grabois, 2004: 42). Similarly, Villamil and De Guerrero (1998) observed the effect of peer revision on writing. They discovered not only that the suggestions made during peer revision were often part of the final drafts the individual learners had produced, but also that "writers made

further revisions on the basis of previous peer collaboration, suggesting a pattern of behaviour conducive to self-regulation." (1998: 508).

It would seem that the metaphors of ZPD and scaffolding, as well as the learner's externalisation of inner thoughts by way of private speech, may help to provide insight in how autonomy can effectively be fostered in foreign language learners. In promoting LA, one of the teacher's roles is to engage learners in the activity of exploring as well as expanding the area of what they can do without the help of teacher or other expert. Learner as well as teacher motivation are seen as decisive factors when LA is promoted. In the final section, we will discuss motivation in the light of the psychological perspectives we have outlined so far.

Just as was the case with humanist and cognitive psychology, constructivism further modifies an LA learning environment, of which the parameters were discussed in the first section of this chapter.

Our discussion of cognitive, humanist and constructivist psychology has provided a sixth and final argument why LA can be viewed as a viable pedagogical construct and goal. Psychologists working within cognitive (Ausubel), humanist (Rogers), constructivist and sociocultural (Vygotsky) frameworks have increased our understanding of the human mind. Together they provide an informed view of how learners learn and of how they may learn more effectively. We will return to these views in the summary at the end of this chapter.

This brings us to the final section of our theoretical chapter on LA. There, we will probe more deeply into the relationship between learner motivation and autonomy in an attempt to establish how the two relate.

3.4 The relationship between motivation and autonomy

As an introduction to how learner motivation and autonomy relate, we will first concentrate on theories on learning motivation. Secondly, we will discuss definitions of motivation within behaviourist, cognitive, humanist and constructivist frameworks. This brings us to a final section, in which we discuss the relationship in more detail.

3.4.1 Theories on learning motivation

It has long been acknowledged that "purely cognitive theories of learning will be rejected unless a role is assigned to affectivity." (Hilgard, 1963:267). Affect refers to a learner's emotions and feelings. Researchers have first of all attempted to bridge the gap between cognition and emotion by a focus on learning styles (Ausubel 1968, Chapelle & Green 1992, Reid 1995, Ehrman 1996, Hoffman 1997, Cohen 1998) and a focus on character types (Myers, 1962; Ehrman & Oxford 1989, 1990, 1995). Learning styles have been defined as "cognitive, affective and physiological traits that are relatively stable indicators of how learners perceive, interact with, and respond to the learning environment" (Keefe, 1979:4). Understanding how human beings feel, respond, believe and value is an exceedingly important aspect of second or foreign language learning (Pike 1967, Arnold 1999, Brown 2000). As a consequence, personality factors and their relation to effective second or foreign language learning have been studied widely. Specific examples of such personality factors are self-esteem, inhibition, risk-taking, anxiety, empathy, extroversion, and introversion (For definitions and overviews of research see Brown, 2000).

The notion of motivation is perhaps the most frequently used term for explaining success or failure in accomplishing complex tasks such as learning another language. Learner autonomy seems to hinge on the learners' willingness and readiness to regulate their own learning. That is why we will consider motivation in more detail in the section below. We will first of all define motivation from behaviourist, humanistic, cognitive and constructivist perspectives. Next, we will explore the relationship between autonomy and motivation in more detail, and

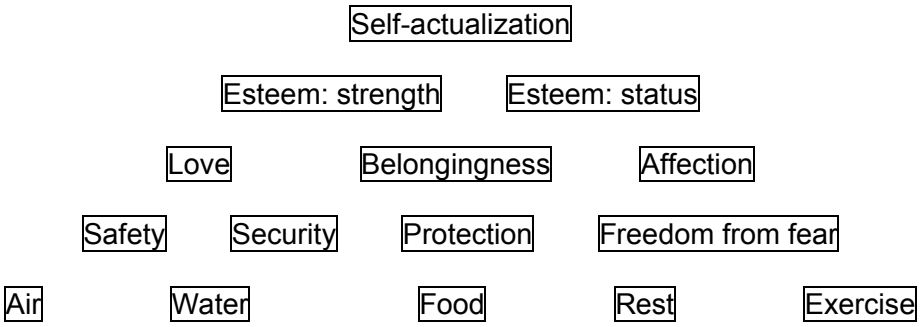
discuss its intrinsic-extrinsic continuum and integrative-instrumental directions. Finally, we will highlight the importance of enhancing intrinsic motivation in the second or foreign language classroom.

3.4.2 Definitions of motivation

The behaviourist, humanistic, cognitive and constructivist theories of mind discussed above all have their own conceptions of motivation. From a behaviourist perspective, motivation is simply considered as a drive to act in the anticipation of an externally administered reward. In Skinner's operant conditioning, a reward serves to positively reinforce desired behaviour. Driven by these reinforcements as well as by previous experiences of reward, human beings act in ways in which others want them to act or behave. Examples of external rewards related to education are praise, gold stars, marks, certificates, diplomas, scholarships, and the anticipation of successful careers, financial independence, and, ultimately, happiness as well. Despite its simplicity, reinforcement theory is a powerful concept in a lot of classrooms. However, a behaviourist conception of motivation offers the learner limited or hardly any room to develop autonomy. Individuals or external forces other than the learner determine how the (s)he should act and behave.

Humanists view motivation from a different perspective. They concentrate on human needs such as affirmation, affection, balance between body and mind, and self-actualization. Maslow (1970) developed his needs theory. It is based on a hierarchical system of human needs that propels us to higher and higher attainment.

table 1: Maslow's hierarchy of needs



Maslow assumes that a person can only move up a level in the hierarchy if a lower foundation has been satisfied. This means that even familiar classroom procedures, such as small talk at the beginning of the class, responding to what the learners have been doing over the weekends, checking homework, helping learners out if necessary, and balancing rest and exercise can pave the way for the learners to meeting higher-order needs. Rogers (1983) did more specifically direct attention to the ways in which the teacher creates a safe and affirmative learning environment and the ways in which (s)he interacts with the learner. The learner's experiences are the starting point for setting a positive climate for learning, helping the learner clarify goals, organizing and making available resources, balancing intellectual and emotional components of learning, and sharing feelings and thoughts with the learner as an equal. All of these elements are considered motivating for the learner.

Cognitive psychological viewpoints offer a less affective and more rationalist approach to motivation. Cognitivists feel the sources of motivation are to be found within the learner, stressing the learner's potential for self-reward. Motivation primarily stems from the learner's decisions, " the choices people make as to what experiences or goals they will approach or avoid, and the degree of effort they will exert in that respect" (Keller 1983:389). Some cognitive psychologists see underlying

drives as the compelling forces behind decisions that human being make. This is e.g. shown in Ausubel's (1968) drive theory. He feels that motivation stems from six basic innate drives, i.e. the needs for *exploration, manipulation, activity, stimulation, knowledge,* and finally *ego enhancement*. According to Ausubel, humans have innate needs to explore, to see 'the other side of the mountain' or to probe for the unknown. They also feel the need to manipulate their environment and cause change. In addition, they need to be active and long for physical and mental movement and exercise. The need to be stimulated by the environment, by other people, or by ideas, thoughts and feelings is also prominently there. All of the previous four needs come together in the need for knowledge, the need to process and internalise the results of exploration, manipulation, activity, and stimulation, to quest for solutions to problems and for self-consistent systems of knowledge (Brown, 2000: 161) Finally, there is the need for ego enhancement, the need for the self to be known and to be accepted and approved by others. Next to Ausubel's drive theory, Hunt's self-control theory (1971) can be seen as another example of a cognitive perspective on motivation. He focuses on the importance of people deciding for themselves what to think, feel or do. It implicitly incorporates a model of autonomy. The assumption is that humans define themselves by making their own decisions, rather than by simply reacting to others. Motivation is highest when one can make choices of one's own, both in short-term and long-term contexts. The learner's basic needs and drives and the need for autonomy can be fulfilled if learners are given the opportunity to make their own choices about what to pursue and what not to pursue, preferably in contexts of cooperative learning. The sociocultural context is what constructivists have elaborated on.

Constructivist as well as sociocultural views of motivation place even further emphasis on the nature of social context. At the same time, these theories considerably elaborate on individual choices (Williams & Burden 1997:120). Each person is motivated differently, and will therefore act on his/her environment in unique ways. Yet, these individual actions and activities are always embedded within a specific milieu and are therefore inseparable from this sociocultural context. Therefore, detailed attention is paid to the didactic realization of core statements such as mediation, interaction, the zone of proximal development, scaffolding, internalisation and inner speech, and private speech. In the next section we will elaborate on the relationship between autonomy and motivation.

3.4.3 Learner autonomy and motivation

To investigate the relationship between learner autonomy and motivation, Dickinson (1995) has reviewed literature about motivation. After describing autonomy as both an attitude towards learning and a capacity for independent learning and using Keller's (1983) definition of motivation as " ...the choices people make as to what experiences or goals they will approach or avoid, and the degree of effort they will exert in that respect" (as cited in Crookes & Schmidt 1991: 389), several theories of motivation are reviewed.

The first one is the extrinsic / intrinsic motivation theory (Deci and Ryan, 1985), which claims that intrinsically motivated learners, i.e. learners who are challenged by learning tasks and interested in their outcomes without any external pressure or situational demand, are much more likely to become effective learners than extrinsically motivated learners, i.e. learners who are mainly motivated by external incentives, such as getting good grades or a certificate. Deci (1975) defined intrinsic motivation and specifically linked the concept to feelings of competence and autonomy.

> *Intrinsically motivated activities are ones for which there is no apparent award except the activity itself. People seem to engage in the activities for their own*

> *sake and not because they lead to an extrinsic reward. ….. Intrinsically motivated behaviors are aimed at bringing about certain internally rewarding consequences, namely, feelings of competence and self-determination.* (1975:23).

Deci's and Deci & Ryan's distinction between intrinsic and extrinsic motivation, should not be mistaken for two orientations originally developed and investigated by Gardner & Lambert (1972). Over a period of twelve years they had extensively studied foreign language learners in Canada, the United States, and the Philippines in an effort to determine how attitudinal and motivational factors determined language learning success. They identified a dichotomy of attitudes that they called "instrumental" and " integrative" motivation. The instrumental side of the dichotomy refers to acquiring a language in order to attain instrumental goals, such as furthering a career or the ability to read technical materials in another language. The integrative side referred to attitudes of learners who wished to integrate themselves into the culture of the second or foreign language group and become involved in social interaction in that group. Originally, Gardner & Lambert referred to the two dichotomies of motivation as types. Later, Gardner & MacIntyre (1991) more appropriately referred to the distinction as two different orientations. That is, depending on whether a learner's context or orientation was (a) academic or career-related (instrumental), (b) socially or culturally oriented (integrative), different needs might be fulfilled in learning a second or foreign language. Within either orientation one can have either high or low motivation. Bailey (1986) illustrated the sets of dichotomies of intrinsic/extrinsic motivation and of instrumental/integrative orientations as follows:

Table 2: Motivational dichotomies

	Intrinsic	**Extrinsic**
Integrative	L2 learner wishes to integrate with the L2 culture (e.g., for immigration or marriage)	Someone else wishes the L2 learner to know the L2 for integrative reasons (e.g., Japanese parents sending kids to Japanese-language schools)
Instrumental	L2 learner wishes to achieve goals utilizing L2 (e.g. for a career)	External power wants L2 learner to learn L2 (e.g., corporation sending a Japanese businessman to the U.S. for language training.)

In Deci & Ryan's theory, intrinsic motivation is a more powerful determinant for learning success than extrinsic motivation. Yet, secondary school contexts are full of extrinsic pressures, such as the school curriculum prescribing what a learner should know and be able to do, parental expectations, society's expectations, an artillery of tests and exams, the habit of teachers and learners to go for immediate gratification and success, expectations that the higher one's level of education is, the more money one will make, competition, and finally the expectations that learners should never fail. According to Deci & Ryan, learner autonomy is enhanced most when learners are given the opportunity to increasingly have a say in what is learned and how this is done.

Another theory reviewed by Dickinson (1995) is attribution theory (Heider, 1958; Weiner 1974, 1980; Ross & Fletcher, 1985). The theory links motivation to the reasons responsible for success or failure in learning. Learners who consider success or failure to be the result of their own efforts to take responsibility are more effective learners than those who attribute success or failure to causes external to themselves or beyond their own abilities. The link between this theory and autonomy is the level of control that learners believe they have over their learning. When they attribute their success or failure to their own efforts, they assume this kind of control and exercise their autonomy as learners.

The motivational model of DeCharms (1984) was added to the three theories described above. In this model, learners are successful when they control their learning and act as "origins", i.e. they originate their own actions. They don't control their own learning when they act as "pawns", i.e. when they exclusively respond to external factors. An example of a successful project based on DeCharms's motivational model is the so-called Carnegie project. It was aimed at enhancing the motivation of low income, black, elementary school children in St. Louis and at reducing their "pawn behaviour".

Dickinson (1995) concludes that there is "substantial evidence from cognitive motivational studies that learning success and enhanced motivation is conditional on learners taking responsibility for their own learning, being able to control and perceiving that their learning successes or failures are to be attributed to their own efforts and strategies rather than to factors outside their control. Each of these conditions is characteristic of learner autonomy as it is described in applied linguistics" (1995: 172).

Van Lier (1996) has also related the concepts of autonomy and motivation. Instead of assuming a dichotomy between intrinsic and extrinsic motivation, he propagates "an interplay between intrinsic (innate) and extrinsic (environmental) factors" (Van Lier 1996: 99). Next, he hypothesizes that in education extrinsic factors tend to dominate to the detriment of intrinsic motivation. Therefore he feels it is necessary to stimulate intrinsic factors, while at the same time taking into account the pressure of extrinsic demands. He advocates seeking a "responsible course of action which balances intrinsic and extrinsic resources and constraints, and the needs and goals of the individual with the needs and goals of society." (Van Lier, 1996: 99). There appears to be a close link between achievement and self-determination in the sense that "… feedback from others can enhance a person's knowledge of success, but only if the person feels that the behaviour was self-determined, and the context was one which facilitated autonomy" (Van Lier 1996: 120).

Dickinson (1995) and Van Lier (1996) have highlighted the strong relationship between motivation and learner autonomy but that, in order to understand that relationship, other variables have to be taken into account if we one is to fathom how the two relate. One important variable is control over learning. If learners can attribute successes and failures to personal efforts instead of factors outside their control, they are more likely to develop autonomy. Another variable is the context of learning. While in education extrinsic motivation is favoured mostly due to e.g. the context of exam and curriculum requirements and time constraints, leading, as is often the case, to "pawn" behaviour (DeCharms, 1984), intrinsic motivation must be stimulated to realise learner responsibility, "origin" behaviour and learner autonomy. The challenge for education is to have learners seek for a balance between the constraints and opportunities of external factors and internal limitations and fortuity, such as the learners' levels and direction of motivation, control over learning and efforts of their own.

This section on the relationship between LA and motivation closes off our first theoretical concern.

3.5 Summary

This chapter focused on what LA is, why it can be seen as a viable pedagogical construct and goal, and what the relationship is between learner motivation and autonomy.

We first addressed the question what LA is. After contrasting and discussing two conceptions of autonomy by Holec (1981) and Kenny (1993), Widdowson (1990) and Little (1991) helped us determine what LA is not. They warned against five misconceptions that were current in views of learner autonomy in the 90s. Then we referred to mainstream definitions other than Holec's and Kenny's. Having discussed these mainstream definitions and interpretations of autonomy, we arrived at thirteen parameters that may help us to understand how autonomy is fostered in learners in formal educational settings.

Then we addressed the question why LA should be seen as a viable pedagogical construct and goal. We arrived at six reasons why a focus on LA in formal educational settings is justified.

Adopting a socio-historical perspective, we arrived at three reasons. The examples of Socrates and Comenius first of all show that we can learn from successful pedagogical contexts in the past, in which learner autonomy was as an important pedagogical goal. Then we asserted that a democratic culture orientation explicitly embraces the notion of self-determination and autonomy. Finally, we claimed that knowledge-based societies require learners who increasingly manage to learn autonomously in the course of their lives.

A fourth reason for an educational focus on LA was derived from adopting a philosophical perspective. We claimed that the reasons-responsive and responsiveness-to-reasoning philosophical accounts of autonomy sensitised us to the importance of increasing a person's motives to act as well as to improve a person's quality of reasoning when autonomy is fostered in formal educational settings.

A fifth justification we mentioned was derived from viewing autonomy from a political perspective. A political perspective invited us to view autonomy in terms of the power dimensions between the more authoritative and expert teacher and the less authoritative and expert learner. Deconstruction and reconstruction of these power dimensions are likely to make learners more autonomous.

A sixth and final justification for fostering LA in adolescent learners, grew from adopting a psychological perspective. We subsequently discussed insights from humanistic psychology, cognitive psychology, constructivism and more in particular from sociocultural theory. These insights may help us arrive at a better understanding of how learners may learn more effectively and how their autonomy as learners is likely to be fostered.

The final question the chapter addressed was about the relationship between learner motivation and autonomy, because LA for a large part seems to hinge on the learner's motivation to learn. We set out with Hilgard's claim that "purely cognitive theories of learning will be rejected unless a role is assigned to affectivity." (1963: 267). We surmised that perhaps the most frequently used term for explaining success or failure in accomplishing complex tasks such as foreign language learning is motivation.

Next, we discussed behaviourist, humanistic, cognitive and constructivist perspectives on motivation.

In an effort to probe more deeply into the relationship between learner motivation and autonomy, we discussed Deci & Ryan's (1985) intrinsic-extrinsic motivational dichotomy and Gardner & Lambert's (1972) instrumental-integrative dichotomy and represented both in a table.

We then referred to attribution theory (Heider, 1958; Weiner, 1994, 1980; Ross & Fletcher, 1985). In this theory, motivation is linked to the reasons learners give for

success or failure in learning. Learners who consider that success or failure is the result of their own efforts to take responsibility are viewed as more effective than those who attribute success or failure to causes external to themselves or beyond their own abilities.

In the motivational model of DeCharms (1984), learners are successful if they control their own learning and act as “origins”, i.e. they originate their own actions. They act as “pawns” if they do not control their own learning and primarily respond to external factors.

Van Lier (1996) also related autonomy to motivation. Instead of assuming a dichotomy between intrinsic and extrinsic motivation, he propagates “an interplay between intrinsic (innate) and extrinsic (environmental) factors” (Van Lier 1996: 99). Van Lier feels that in education extrinsic factors tend to dominate to the detriment of intrinsic motivation. Therefore he feels it is necessary to stimulate intrinsic factors, while at the same time taking into account the pressure of extrinsic demands.

Following Dickinson (1995) and Van Lier (1996), we emphasised the strong relationship between motivation and learner autonomy and supported the plea for more attention to intrinsic motivation and self-regulated learning in formal education.

CHAPTER 4: COMMUNICATIVE COMPETENCE IN FOREIGN LANGUAGE EDUCATION

4.1 Introduction

In the previous chapter we discussed what *learner autonomy* involved and why it should be considered as a relevant pedagogical goal. We particularly highlighted the relationship between learner autonomy and motivation. In this chapter we will focus on the construct of *communicative competence* in foreign language education and explore some of its backgrounds.

Communicative competence is the common name of the knowledge and skills allegedly needed to communicate correctly, efficiently, and appropriately. Communication has come to be seen as the ultimate goal of foreign or second language learning. Languages are learned with the aim to use them and communicative language education (CLE) has come to be associated with progressive, innovative teaching, except by those who see it as an appeal to an "anything-goes-as-long-as-you-get-your-meaning-across" approach to foreign language learning. (Savignon, 1997:7). Such opponents of communicative approaches are of opinion that efforts at meaningful communication can only be successful if they are based on a profound grammatical and idiomatic foundation. We will discuss communicative competence in foreign language education in four sections.

In specifying what communicative competence is, we will first explore some of its backgrounds. Our first exploration is once again historical. Starting with a concise methodological history of second and foreign language education, we will see how an almost exclusive focus on form contrasts with approaches that focus on meaning and language use to varying degrees.

Next we will move towards a definition of communication and the competences required to effectively learn how to communicate in a second or foreign language. This will be done by first addressing two theoretical issues in language education, i.e. language as social behaviour and language functions.

After that we will discuss representatives of three main interpretations of the construct of communicative competence identified by Savignon (1983,1997), i.e*. from surface structure to meaning*, *from meaning to surface structure*, and *specification of context*.

In a final section, we will go into some research interests and findings, which are characterised respectively by a focus on meaning, form and the cultural aspects of language use.

Similar to our approach in the chapter on learner autonomy, the present chapter aims to provide us with a theoretical framework that can be used to generalise the findings of this multiple-case study to how communicative competence is viewed in theory.

4.2 A methodological history of second and foreign language education

Similar to our discussion of LA, we will approach communicative language education and its central construct *communicative competence* from an historical perspective.

Mackey (1965) was among the first to document the history of foreign language teaching in a systematic way. The summary of his final conclusions has been quoted frequently when foreign language teaching (FLT) is considered from a historical perspective.

> *While sciences have advanced by approximation in which each new stage results from an improvement, not a rejection, of what has gone before, language-teaching methods have followed the pendulum of fashion from one extreme to the other.* (Mackey 1965:138)

Mackey very much regretted the apparent lack of progress after so many centuries of development of FLT. This was first of all caused by the fact that 'penduli of fashion' had not always been studied in a scientific way. The growing awareness of 'development without progress' gave rise to a number of publications and studies in the field of historical FLT (Kelly (1969), Schröder (1980), Howatt (1984), Howatt & Widdowson (2004) and Wilhelm (2005)). Wilhelm, who explored the history of English and foreign language teaching in the Netherlands from 1800 – 1920, holds a convincing plea for studies documenting the history of FLT.

> *Not until we relate our knowledge of practice and theory in historical FLT to our own ideas, do we arrive at a better understanding of present-day FLT. And not until that moment will there be real progress.* (Wilhelm, 2005:22)

The crucial question is what foreign language practitioners can learn from the past. What can be understood about *what* has been taught, *how* this has been done, and *why* a foreign language has been taught in a particular way?

Van Els, T.; Bongaerts, T.; Extra, G.; van Os, C. & Janssen-van Dieten, A. (1984) offered an additional argument why FLT had developed in different directions without clearly traceable progression. A first problem they signalled was a wide number of interpretations of the term 'language-teaching method'. Van Els et al. follow Schröder & Weller (1975) by claiming that the word 'method' has been used rather haphazardly. 'Methods' were "sometimes named after their 'inventor', occasionally after their objective, at times after their didactic principles, and now and again after the aids employed…"(1984:144). The term method was also used in confusing ways outside FLT. Method could also refer to the different ways in which teaching was organized, such as individualized versus group teaching. It was also used for different principles of gradation of course content, as shown in differentiation or programmed instruction. 'Method' could also refer to particular ways of grading course materials, as in linear versus cyclic gradation. The term has also been used for different types of classroom methodology, as in labels like 'lecturing method' or 'self-instructional method'. Even different types of exercise were sometimes referred to as method, as in the 'pattern-drill method'. To finally add to the confusion, it is not unusual to refer to a particular course book as a 'method'. The term 'method' in this study will invariably be used as defined as a broad overall approach by Van Els et al.

> *What is meant here by the word 'method' is the total of considerations concerning the specification of objectives, the selection and gradation of course content, and the selection of didactic procedures. Put differently, what we mean by 'method' is the coherent whole of all considerations concerning what is taught and how it is taught.* (1984:144)

Before turning to an overview of FL-methods, an important statement has to be made. It is essential to realise that empirical data so far has not provided support for the myth that there could be one single most effective approach or 'method' to second and foreign language learning. (Jarvis 1972, Van Els et al. 1984, Brown 2001). Language learning and teaching involve complex processes with a huge number of variables that cannot be simplified into a "one-and-only" approach that is effective for each and every learner in each and every sociocultural context. What applied linguists and educational practitioners can at best strive for is to be first of all well-informed about the latest findings and views in source disciplines as varied as

linguistics, pedagogy, educational science, sociology and psychology. Secondly, researchers and teachers have to realize that data, insights and achievements from source disciplines can rarely ever be adopted directly without any adaptations. What they can, then, at best hope for is to marry ideas and suggestions in such a way, that most of the learners and teachers perceive positive change as compared to the ways in which they used to learn or teach. Such an eclectic view had already been advocated by Jespersen (1904, 1967:14), when he claimed that the new Reform Method '…is not the whim of one man, but the sum of all the best linguistic and pedagogical ideas of our time, which, coming from different sources, have found each other, and have made a beautiful alliance for the purpose of overturning the old routine.' There is still no single most effective 'method' to second and foreign language learning for the simple fact that learning another language involves complex processes with a huge number of context-dependent variables. Nevertheless, a number of trends can be discerned that were favoured at given periods in time. Our perspective will be socio-historical once more, because the trends were very much linked to the demands of foreign language knowledge and skills.

Van Els, et al. report that the nineteenth and twentieth centuries show an increasing demand for foreign languages. In earlier days in Europe, when only few learners sought to learn foreign or second languages, the usual procedure was to hire a private tutor. In this way many a young Roman was educated bilingually in Latin and Greek from a very early age. In the Middle Ages, when Latin was still a means of communication, the language was taught in an intensive and direct way, that is the language was used as a means of instruction for all subjects from the start (1984:146). In Renaissance days, it became the craze for the fortunate few to visit the country in question to learn the language and explore its culture.

In the course of the nineteenth and twentieth centuries, a number of approaches were used or developed, which shared more or less the same characteristics. It is important to stress our list of popular methods is neither exhaustive nor sequential. We will not attempt to list all possible methods that have ever been referred to as such. Neither does the list represent a chain in which one particular method was simply superseded by another in history. The methods should best be seen as trends that were popular at a particular time and place. Even present-day curricula may show combinations of trends or methods that used to be popular before.

In our discussion of popular methods in foreign language teaching, we have singled out six. The methods are particularly relevant for understanding the European context of FLT, and more in particular the context of English as a foreign language in the Netherlands around the new millennium. We will subsequently discuss the grammar-translation method, the direct method, the reform method, the audiolingual method, and finally the related audiovisual method, with one of its most notable representatives: the structuro-global audiovisual method (SGAV). We will end the section by contrasting the loss of popularity of audiolingualism with the advent of more communicative approaches to second and foreign language learning.

4.2.1 The grammar-translation method

This method is often referred to as 'traditional' in the sense that it reflects the way in which Latin and Greek had been taught for centuries. Mastery of Latin and Greek helped learners study the classics and train their minds. Education was characterized by logical analysis of the language, extensive memorization of complicated rules or paradigms, and the application of these rules and paradigms in translation exercises (Rivers 1968:15). Adopting this approach to foreign languages gave them a status comparable to Latin and Greek. This was particularly the case in the nineteenth century, when foreign languages actually had to compete with the classical languages.

The grammar-translation method heavily relies on rote-learning, both of grammatical rules or paradigms and of extensive bilingual vocabulary lists. Its most popular exercise is translation of, sometimes contrived and artificial, sentences or texts from one language into the other. Oral skills are neglected, the predominant language of instruction is L1, and little or no attention is paid to pronunciation practice. Characteristics of the method are still to be found in foreign language curricula.

The grammar-translation method with its dissociation from authentic language and L2-use is furthest removed from a communicative approach of learning another language. However, this does not necessarily mean that translation exercises from one language into another cannot be helpful in understanding how languages are structured and which words are used to express meaning most appropriately. We will return to this issue when we discuss recent research interests in linguistic structure as a means to better understand and enhance communicative competence.

4.2.2 The direct method

Rather than a single method, the direct method is a collection of approaches and techniques. That is the reason why the direct method rarely ever exists in its purest form. The collection of direct approaches and techniques were a reaction against the objectives and procedures of the grammar-translation method. The oral skills, which had been underrated in grammar-translation approaches, are very much emphasized. The L2 is used intensively and preferably exclusively. All of the L2 utterances are meant to be meaningful in the sense that they are directly associated with the objects and actions they refer to. Listening is thought to be best learned by listening practice, speaking is best learned by speaking practice. Grammatical rules are not explicitly formulated, or at least not explicitly taught. The learner acquires 'knowledge' of grammatical structures inductively by practising with complete and meaningful utterances (Rivers 1968:18ff.).

Berlitz, a German who emigrated to America in the nineteenth century, founded a chain of schools in which the direct method was used in a pure form. Right from the start, L2 is the only language used. By way of question-and-answer dialogues between the teacher and the learner, skills in L2 are gradually developed. The schools are still popular in many cities around the globe.

The direct method contrasts with the grammar-translation method to the full. Oral skills are emphasised and L2 is used intensively in meaningful ways. Learners learn to listen by listening, and to speak by speaking, building up their own grammar and idioms as they go along. As such, the direct method very much holds the promise of improving a learner's communicative competence. Yet, a direct method is not free from criticism. One question in particular begs to be answered. In how far are learners of a foreign language who are taught two or three classes a week and who are not subjected to the foreign language outside school capable of building up their own grammar and idioms? As we will see later in the chapter, models of communicative competence all include linguistic competence as an essential part. This strongly suggests that it is a structural knowledge and skills component that cannot be discarded.

4.2.3 The reform method

Four innovative linguists are associated with more theoretically underpinned versions of the direct method that are commonly referred to as the reform method. The linguists are Wilhelm Viëtor (1850-1918), Otto Jespersen (1860-1943), Henry Sweet (1845-1912) and Harold E. Palmer (1877-1949). All of them shared an interest voiced in Viëtor's influential pamphlet *Der Sprachunterricht muss umkehren*, published in 1882 under the pseudonym of 'Quousque Tandem' (=For how long will this be going on). Viëtor stressed the importance of teaching language that the

learner can use in order to communicate in day-to-day situations. The language taught and learned should not primarily and exclusively be literary language. The linguists, pedagogues and teachers who supported the reform method demanded a focus on 'living language'. All of the four innovative linguists had their own unique contributions to the reform method.

Sweet, e.g. emphasized the teaching of pronunciation based on knowledge from the field of phonetics. Besides, he felt that every language had its own structure, and could therefore not be forced into the straitjacket of Latin grammar. He was not in principle against the use of translation and explicit rules of grammar. Another instance of Sweet's main merits is that he developed a theoretical model for FLT in which a distinction was made between objectives, selection of course content, and gradation of course content.

According to Darian (1969:546), Jespersen paid a lot of attention to didactic principles in his work, such as the use of contextual materials and of drills in which the structural pattern was closely related to meaningful communication. Jespersen strongly opposed practice with monotonous and uninteresting materials or disconnected sentences.

Palmer was a linguist who advocated and searched for a scientific foundation for FLT. Palmer's approach was eclectic in the sense that he wished to base his study of FLT on an integration of principles and data from linguistics, pedagogy, and psychology (Titone 1968:60). He also was less prejudiced than others towards traditional teaching, and had always been keen neither to wholly reject the past, nor to uncritically accept the future. As an example he called the exclusion of translation as a means to clarify meaning 'an uneconomical and unnatural principle'(Palmer 1917:1968:60).

The representatives of the reform method were also critical of the ubiquitous grammar-translation method. Their criticism was primarily based on the theories and disciplines that had shaped them as scholars. The result was that they all attempted to make language learning more meaningful, without losing sight of recent developments in linguistics and without wholly rejecting the past. Viëtor's focus on 'living language' in day-to-day situations, Sweet's attention to the role of pronunciation, Jespersen's use of contextual materials and drills related to meaningful communication, and Palmer's eclecticism transcending the discipline of linguistics, are all examples of approaches that potentially help learners to become better users of another language.

4.2.4 The audiolingual method

Audiolingualism attempted to offer an alternative to the grammar-translation method as well. Audiolingual methodology was supported by theory, i.e. behaviourism and structural linguistics by e.g. Fries (1945), who had explored differences between languages in sound systems, structures, and vocabulary. The methodology was elaborated in the Army Specialized Training Program (***ASTP***) in the 1940s, which arose from the need to train soldiers in second and/or foreign languages. An important didactic principle of the audiolingual method used in the ASTP was captured by Bloomfield (1942:12): '*Language learning is overlearning: anything less is of no use.*' For decades to come, course materials were designed and used in which new language material was presented in the form of dialogues with a central role reserved for repetition and memorization. Fries (1945,1952), Lado (1957), and Brooks (1964) further developed the use of pattern drills, and supported using the most modern of auditory aids: the language laboratory. Contrastive analysis and pattern drills were characteristic of an audiolingual curriculum. There was hardly any explanation of grammar. Audiotapes and visual support were frequently used. Starting with the Army Specialized Training Programs, the American 'Golden Age' of audiolingualism was from 1958 to about 1964. In

Europe it was a popular method in the late 60s and 70s, too. However, audiolingualism never assumed a position any way near the monopoly it has held for so long in the US (Strevens 1972: 717).

The audiolingual method was informed by behaviourism, structural linguistics and the increasing availability of audio-technology. This resulted in stimulus-response pattern drills, contrastive analysis of the languages involved, and the use of language laboratories. With regard to language use, the audiolingual method is just a little less far removed from communicative approaches than the grammar-translation method is. Its general adage that learning is overlearning was exploited to an often boring full. Yet again, this does not mean that aspects of audiolingualism, such as contrastive analysis and the use of audio-input and learner output, necessarily hinder the development of communicative competence. After all the Army Specialized Training Programs succeeded in teaching the soldiers some linguistic survival skills.

An interesting conglomeration of approaches that represents a more European methodological perspective on foreign language teaching is the audiovisual method.

4.2.5 The audiovisual method

Approaches referred to as 'audiovisual' all share a belief in a language learning context full of audio- and visual stimuli. Rivers (1968:175) and Titone (1968:108) tone down the role of these stimuli and see the added visual components as no more than a collection of useful aids and techniques that can be used in various methods. Van Els et al. (1984: 153) attach more importance to audiovisualism and discuss the approaches as a method in its own right. To them visual elements and realia, such as slides, pictures and films, may provide an effective learning context both in the presentation and exercise phases. They discuss a relatively influential instance of an audiovisual method: the so-called structuroglobal audiovisual method (SGAV).

The method was developed first by the Centre de Recherché et d'Etudes pour la Diffusion du Français at Saint-Cloud in France. Just as the audiolingual methods, SGAV was oriented towards structuralism. However, Van Vlasselaer (1972:26) had called American structuralism a 'structuralisme dèsincarné'. He was looking for an interpretation of structuralism that looked ahead at sociolinguistic and pragmalinguistic theories that would be elaborated later (Van Els, 1984:154). This is one of the reasons why Guberina (1972:26) felt the need to add the word 'global'. The addition was meant to indicate that every structure should be viewed as embedded in a situation of language use. Thus, the didactic procedures of SGAV should enable that:

1. every linguistic structure dealt with is embedded in a situation of language use;
2. visual components are integrated, so that the originating situational context provides a starting-point for verbal response and helps the learner dissociate herself from L1;
3. a great deal of early emphasis is on pronunciation, not just of isolated sounds but particularly of suprasegmental and prosodic elements.

In more than one respect this typically European 'method' is a prelude to the communicative approaches that were to be developed later, with its linguistic structures embedded in language use, integrated visual components, and attention to functional pronunciation.

Our methodological excursion so far has shown the importance of being eclectic in the sense that the past is neither neglected, nor the future of communicative approaches uncritically accepted. Many aspects of the methods we have discussed may potentially help to develop a learner's communicative competence. Criticism of the more communicative approaches seem to centre around the roles that grammar

and idioms, and in its broader sense linguistic competence, play when learners learn to communicate in another language.

We will end our methodological overview by contrasting audiolingualism with more communicative approaches and offer four reasons why audiolingualism lost its popularity when second or foreign languages are taught and learned.

4.2.6 Audiolingualism vs. more communicative approaches

The notion that language is a means of communication and should be taught and learnt as such, has become more and more accepted. As we have argued above, discussions of communicative approaches focus on the role of linguistic competence in relation to other aspects and/or competences that allegedly make up communicative ability. As early as 1968, Jakobovits pointed out the need to be more explicit about the role of linguistic competence. He talked about what it is to know a language.

> *Everyone accepts the notion that language is a means of communication, but there is much less agreement about just what is involved in the ability to communicate. The distinction between "linguistic competence" and "communicative competence" is either not explicitly taken into account in the majority of FL courses or it is tacitly assumed that the former must precede the latter in such a way that a certain high level of linguistic competence must be achieved before attempting the functional use of the FL.* (1968, p.184)

Jakobovits's words almost seem prophetic. For decades to come practitioners would struggle to define communicative competence and specify the role of linguistic competence in acquiring communicative skills in a second or foreign language. This struggle is shown by the persistent influence of both grammar-translation and audiolingualism, long after its initial popularity has decreased.

The audiolingual method in particular is often compared and contrasted with more communicative approaches to second and foreign language learning. The attention for communicative language education can be seen as a response to the rather mechanical way in which languages were taught and learned in audiolingual methodology. Savignon compares and contrasts six basic tenets of audiolingual methodology with six guiding characteristics of a communicative approach to second and foreign language teaching. (1997: 17/8 & 25/6)

Audiolingual methodology	Communicative approaches
1. Language learning is habit formation.	1. Language use is creative. Learners use whatever knowledge they have of a language system to express their meaning in an infinite variety of ways.
2. Language performance consists of four basic skills: listening, speaking, reading and writing.	2. Language use consists of many abilities in a broad communicative framework. The nature of the particular abilities needed is dependent on the roles of the participants, the situation, and the goal of the interaction.
3. L2 learning, like L1 learning, should begin with listening and speaking, regardless of the end goal of the learner.	3. L2 learning, like L1 learning, begins with the needs and interests of the learner.
4. A contrastive analysis of the phonological and structural differences between L1 and L2 provides the most effective basis for materials development and sequence.	4. An analysis of learner needs and interests provides the most effective basis for materials development.
5. The basic unit of practice should always be a complete structure. Production should precede from repetition to substitution and continue until responses are automatic. Spontaneous expression should be delayed until the more advanced levels of instruction. Production errors in structural or phonological features mean that the patterns have not received sufficient prior drilling.	5. The *basic* unit of practice should always be a text or a chunk of *discourse*. Production should begin with the conveyance of meaning. Formal accuracy in the beginning stages should be neither required nor expected.
6. The teacher is the centre of all classroom activity and is responsible for maintaining attention and a lively pace. (Savignon. 1997, pp. 25/6)	6. The teacher assumes a variety of roles to permit learner participation in a wide range of communicative situations. (Savignon, 1997, pp. 28/9)

The table above invites two comments. Interestingly, we see that learner needs and interests and the teacher permitting learner participation strongly suggest the inclusion of a learner autonomy model in communicative approaches. Secondly, the table above may suggest that audiolingual and communicative methodologies have little in common and almost seem mutually exclusive. In educational practice, however, mixed forms are found. The mainly narrative chapters 7 to10 will show in how far this is the case for the three teacher respondents in this study.

We will end our methodological focus by reasoning why audiolingualism lost its original appeal. Four reasons come to mind why audiolingualism partly lost its initial popularity. First, its approach did not automatically lead to more effective and correct language use if the learners had to do without the usual audiolingual classroom stimuli and correct responses. A lot of critics felt that learners did not build up relevant knowledge of the formal system of the language and were unable to understand and use a wide range of vocabulary. In addition, the method did not at all take learner needs, preferences, and views into account.

There is a second argument for the loss of popularity of audiolingual methodology. Behaviourism was gradually replaced by a host of cognitive and constructivist learning theories. Stimulus-response patterns and positive reinforcement of desired behaviour were no longer seen as the main characteristics of an environment conducive to learning. The social-interactionist and developmental

character of learning was taken into account more and more, as we have discussed in the previous chapter on learner autonomy.

A third reason why the audiolingual method lost some of its popularity was its focus on cognitively-oriented aptitude tests. From the early 1970s onwards, assessment had become more competence-oriented. Today, competence-oriented assessment has come to be widely embraced when professionals are educated. Yet, this recent focus on assessing and evaluating competences is far from new. McClelland (1973), a Harvard psychologist, received wide acclaim in the business world for his work in defining and assessing competences for a particular job. (Savignon, 1997:8). He was critical of the use of cognitively-oriented aptitude tests, and advocated a more meaningful approach to assessing the competences required for a particular job. He preferred to construct and administer tests of skills that were based on proficiency on the job. “If you want to test who will be a good policeman, go find out what a policeman does. Follow him around, make a list of his activities, and sample from that list in screening applicants.”(McClelland, 1973: 7). He advised test constructors to leave their desks and their “endless word and pencil and paper games” to engage in the formidable task of sampling valid criterion behaviour. Once these criterion behaviours have been established, they can be made explicit to those in preparation of a required performance test. McClelland stressed the importance of making explicit to the learner what particular criterion behaviour was going to be tested. The psychologist, teacher, and learner may then openly discuss the applicant’s performance on directly observable, task-related skills.

A final reason why audiolingualism lost its place as a leading methodological approach in foreign and second language is that the concept of communicative competence gradually became embraced. Communicative competence was defined as a feasible construct for second and foreign language education due to the works of sociologist and anthropologist Hymes (1967, 1971,1974), and the works of applied linguist Halliday (1973, 1978). These insights were united in a model of communicative competence developed by Canale & Swain (1980), which was refined by Canale (1983) and later modified by Bachman (1990).

In Europe in the 1970s, a parallel and specifically European interest in foreign language education was developing. The Council of Europe had united pragmalinguists of the various member states, such as Alexander, Coste, Van Ek, Richterich, Roulet, and Wilkins, to work on curricula and materials that focused on specification of objectives and immediate practical communication in another language. It led to the development of notional-functional syllabuses, the Threshold level, and a ‘unit-credit’ credit system that enabled learners to study ‘units’ of work in specific notional-functional areas and gain ‘credits’ for these, and finally to compiling the ‘Common European Framework of Reference for Languages: Learning, Teaching, and Assessment’ (Council of Europe, 1998, 2001). This comprehensive description of language use, the competences involved, and descriptors to assess these competences will be discussed as an interpretation of communicative competence Savignon (1983,1997) labelled as *specification of context*.

In the next section, we will move towards a definition of what communicative competence is. The focus will first be on some theoretical issues that helped to define what knowledge and skills a learner needs to be able to communicate in a second or foreign language, i.e. language as social behaviour and language functions.

4.3 Towards a definition of communicative competence

Savignon (1997) explains that the course of language teaching methodology, which includes the development of communicative language teaching (CLT), has

never run smoothly. The concept of CLT has been informed by theoretical developments in disciplines as varied as linguistics, psychology, sociology, and anthropology. CLT has also been affected by particular socio-political contexts. Pressures in the market place, on the political front, and in society at large called for the need to develop communicative language abilities in *human capital*, as language learners are sometimes ungraciously called in the knowledge societies towards which we are moving. These pressures affected foreign language approaches, methods and goals. One of the issues Savignon raises here is that "Teaching always has been, and always will be as much art as it is science." (1997:15). She adds that this "should not deter us from elaborating methods, systematically trying them out, and judging the results. The theoretical support for what may or may not work in practice is of interest insofar as it provides a broad view of the directions we are pursuing."(1997:16). In her study Savignon attempts to reconcile the pedagogical needs and concerns of language practitioners with discussions and insights from psychology, linguistics, and communication theory. The theoretical issues she addresses are the view that language is increasingly seen as *social behaviour* and that language use can be characterised by a variety of *functions.*

4.3.1 Language as social behaviour

At the beginning of the 20th century, structural linguists like Bloomfield (1933) had been concerned with procedures for isolating phonemes and morphemes in linguistic descriptions. In the mid-twentieth century Chomsky started to object to this exclusive focus on "surface" features of phonology and morphology and concerned himself with "deep" semantic structures . Transformational-generative grammarians like Chomsky assume there is an underlying grammatical competence common to all native speakers. This underlying grammatical competence is manifested in language performance. Chomsky was primarily interested in "…an ideal native speaker-listener, in a completely homogenuous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitation, distractions, shifts of attention and interest, errors (random or characteristic) in applying his knowledge of the language in actual performance." (Chomsky, 1965:3).

This view of an "ideal speaker-listener" as a nonexistent abstraction was challenged by linguist and anthropologist Hymes. He looked at the real speaker-listener in a feature of language that was outside Chomsky's scope, i.e. social interaction. Hymes felt that members of a community will behave and interpret the behaviour of others according to the knowledge of the communicative systems available to them. This knowledge includes the formal aspects of language, such as grammar. However, structural knowledge was not the only parameter to the systems of rules that underlie communicative behaviour. He proposed four parameters:

1. *Whether (and to what extent) something is formally possible;*
2. *Whether (and to what extent) something is feasible;*
3. *Whether (and to what extent) something is appropriate (adequate, happy, successful) in relation to a context in which it is used and evaluated;*
4. *Whether (and to what extent) something is in fact done, actually performed, and what its doing entails.* (Hymes, 1971:12)

Hymes' concept of competence included both tacit *knowledge* of the four parameters outlined above and *ability for use.* Ability for use relates to all four parameters and includes non-cognitive factors such as motivation, attitude, and interactional competence, i.e. composure, courage, and sportsmanship (Savignon, 1997:18). The actual performance of a person is the result of the interaction between the person's competence, the competence of others, and the nature of the event itself as it unfolds.

Hymes (1967, 1971, 1974) added a social-interactionist aspect to the theory of communicative competence. Another perspective was added by Halliday (1973, 1978).

4.3.2 Functions of Language

Halliday concentrated on the functions of language. He was of opinion that only by looking at language in use or in its *context of situation* are we able to understand the functions served by a particular grammatical structure. He objected to the distinction Hymes made between competence and performance. Halliday found the distinction "unnecessary if it is just another name for the distinction between what we have been able to describe in the grammar and what we have not, and misleading in any other interpretation.." (Halliday, 1970:145).

Halliday used the term function to mean the purposive nature of communication. He first distinguished three basic functions (1970), which he later expanded to seven microfunctions (1973). The three basic functions he identified were the *ideational, interpersonal,* and *textual* function of language use.

> 1. *Language serves for the expression of 'content': that is, of the speaker's experience of the real world, including the inner world of his own consciousness. We may call this the* ***ideational*** *function. … In serving this function, language also gives structure to experience and helps to determine our way of looking at things, so that it requires some intellectual effort to see them in any other way than which our language suggests to us.*
> 2. *Language serves to establish and maintain social relations…through this function, which we may refer to as* ***interpersonal****, social groups are delimited, and the individual is identified and reinforced, since by enabling him to interact with others language also serves in the expression and development of his own personality.*
> 3. *Finally, language has to provide for making links with itself and with features of the situation in which it is used. We may call this the* ***textual*** *function, since this is what enables the speaker or writer to 'construct texts, or connected passages of discourse that is situationally relevant; and enables the listener or reader to distinguish a text from a random set of sentences.*
> (Halliday, 1970: 143)

The seven microfunctions that Halliday identified later (1973) were respectively the *instrumental*, *regulatory*, *representational*, *interactional*, *personal*, *heuristic* and *imaginative* functions to classify purpose in communication.

Brown (2000: 251) offers the following examples. The *instrumental* function of language serves to cause things to happen, as in sentences like "I want you to leave" or "This court finds you guilty". The *regulatory* function is the maintenance of control of certain events, as in "If you behave, you won't regret it." or " Keep up doing the good work.". Approval, disapproval, behaviour control, setting laws and rules are all regulatory functions of language. The *representational* function is the use of language to "represent" reality as one sees it, with examples like "The sun is hot." or "John seemed rather absent-minded at the meeting last night.". The *interactional* function serves to establish social contact and keep channels of communication open. Successful interactional communication requires knowledge of slang, jargon, jokes, folklore, cultural mores, politeness, and formality expectations, and other keys to social exchange. The *personal* function allows a speaker to express personal aspects of his/her individuality, such as feelings, emotions, or "gut-level" response. In the personal nature of language, cognition, affect and culture all interact.

The *heuristic* function refers to the language used to acquire knowledge or to learn about the environment in its broadest possible sense. A characteristic example of the heuristic function is young children's incessant "why" questions about the world around them, by which they elicit representations of others. Finally, the *imaginative* function allows a person to create ideas or imaginary systems that go beyond the reality of the real world, as in creating impossible dreams, writing prose, poetry or lyrics, telling jokes, in short any means that allows one to soar the heights of the beauty of language.

The seven functions of language are neither discrete nor mutually exclusive. A single sentence or conversation might incorporate several functions at the same time. Language functions and their interpretations by others are largely determined by the social context. Yet, the crux of foreign and second language learning is more often than not on linguistic form. A learner might acquire correct word order, syntax and a fair amount of lexis, but still not understand how to achieve a desired and intended function through careful selection of the words, intonation, nonverbal signals, and astute perception of the context of a particular stretch of discourse. (Brown, 2000: 252).

In the next sections we will attempt to more specifically define what communicative competence is. We will do so by first discussing two influential models of communicative competence. Next, we will go into Savignon's constructivist definition of communication, which subsequently leads to five statements related to communicative competence.

4.3.3 Communicative competence

Both Hymes's focus on language as social behaviour and Halliday's attention to language functions combined in two models of communicative competence that have been influential over the years. The first model is Canale & Swain's (1980) and the second model is Bachman's (1990), which was largely based on Canale & Swain's and Halliday's insights.

Canale & Swain's model suggests four components of communicative competence. The first two underlying competences relate to the use of the linguistic system itself and the last two competences define the functional aspects of communication.

- **Grammatical competence**:
  Knowledge of the formal structures of a language, which involves the ability to recognize its *lexical, morphological, syntactic* and *phonological* features and to manipulate these features to form words and sentences. A person demonstrates grammatical competence by using a rule, not by stating a rule.

- **Discourse competence**:
  Knowledge of the ways in which sentences are connected in a language to form a meaningful whole, which involves the ability to interpret a series of sentences or utterances in order to form a meaningful whole and to achieve coherent texts relevant to a given context.

- **Sociolinguistic competence**:
  Knowledge of the social rules of language use, which involves the ability to appropriately use language in a given context, taking into account the roles of the participants, the setting, and the purpose of the interaction.

- **Strategic competence**:
  Knowledge about how to compensate for imperfect knowledge of linguistic, sociolinguistic and discourse rules, which involves the ability to cope within the limitations experienced and/or deal with limiting factors such as fatigue, distraction or inattention.

The four components represent neither hierarchy nor sequence. It is open to speculation how the four components interact in a given situation. All of the components are interdependent, and none of the components is greater than the whole of communicative competence.

Canale & Swain's model has undergone some modifications over the years. These modifications are perhaps best captured in Bachman's (1990) model of communicative competence. Bachman reorganized and extended Canale and Swain's model of communicative competence. He saw communicative language ability as made up of five components:

- knowledge structures (knowledge of the world)
- language competence (knowledge of language)
- strategic competence (the ability to assess, plan and execute appropriate interactional language use)
- psychophysiological mechanisms (the neurological and psychological processes involved in producing and comprehending language)
- context of situation.

Language competence itself was divided into the broad headings of *organizational competence*, which includes both grammatical and discourse (or textual) competence, and *pragmatic competence*, which includes both illocutionary (i.e. functional) competence and sociolinguistic competence.

Canale & Swain's and Bachman's models help to define communicative competence. Whichever model one prefers, communicative competence is made up of a linguistic or organizational component, a pragmatic component and a strategic component. In Bachman's model, language competence, which is subdivided into organizational competence and pragmatic competence, is related to general knowledge of the world, to psychophysiological mechanisms and to the specific contexts in which language is used.

With communicative competence thus broadly defined, it is useful to consider in more detail how Savignon defined communication. After all, in acts of communication the competences identified so far interrelate in often complex ways. Savignon attempted to unite education, teacher education, and research when she defined how she saw communication.

> *Communication, then, is a continuous process of expression, interpretation, and negotiation of meaning. The opportunities for communication are infinite and include systems of signs and symbols (which we cannot begin to classify or even identify), of which language is but a part. The color of our skin, the way we dress, the way we wear our hair, the way we stand, smile, listen, nod, and pause all communicate information to others along with the sound of our voice and the words we speak. We are concerned with communication from birth, and we learn to respond in new contexts as we accumulate life experiences. However, the meaning we intend and the meaning we convey are often not the same. In going from thoughts and feelings to their symbolic representation-in written or spoken words, gesture, design, color, movement, or sound- choices must be made. We make the best use we can of the symbolic systems we know. The meaning we convey depends on others who*

> *share an understanding of these symbols and who may or may not interpret them as we intend.*
> (1997:14)

Savignon views communication as a continuous process of expression, interpretation and negotiation of meaning. Constructivist perspectives resound in this apt description of communication. In line with this specification, Savignon (1997: 14/5) sums up five statements related to communicative competence.

> 1. *Communicative competence is a dynamic rather than a static concept. It depends on the negotiation of meaning between two or more people who share to some degree the same symbolic system. In this sense, then, communicative competence can be said to be an interpersonal rather than an intrapersonal trait.*
> 2. *Communicative competence applies to both written and spoken language, as well as to many other symbolic systems (i.e. facial expressions, gesture, design, color, movement, and sound)*
> 3. *Communicative competence is context specific. Communication takes place in an infinite variety of situations, and success in a particular role depends on one's understanding of the context and on prior experience of a similar kind. Success requires making appropriate choices of register and style in terms of the situation and the other participants.*
> 4. *There is a theoretical difference between competence and performance. Competence is defined as a presumed underlying ability and performance as the overt manifestation of that ability. Competence is what one knows. Performance is what one does. However, only performance is observable, and it is only through performance that competence can be developed, maintained and evaluated.*
> 5. *Communicative competence is relative, not absolute and depends on the cooperation of all the participants. It makes sense, then, to speak of degrees of communicative competence.*
> (Savignon, 1997: 14/5)

Savignon's definition of communication and five related statements show that communication may often involve complex processes that can be interpreted from a variety of angles. An example of such an interpretation is the role of linguistic competence in communication, i.e. the role of alleged building blocks as grammar, vocabulary, and pronunciation and the role of discourse competence. Savignon (1983,1997) identified three main interpretations of communicative competence. The three interpretations help to partly unravel the complexity of communicative competence as a generally accepted construct and pedagogical goal.

4.4 Interpretations of Communicative Competence:

Savignon (1983,1997) has identified and labelled three mainstream interpretations of communicative competence: *from surface structure to meaning*, *from meaning to surface structure*, and *specification of context*. The three perspectives on communicative competence and some of its representatives will be discussed in more detail below.

4.4.1 From surface structure to meaning

Savignon mentions three representatives of the view that classroom goals of communicative competence will only be met if learners are provided with systematic practice in the use of structures and vocabulary that have previously been introduced. Paulston (1974), Rivers (1972), and Valette (1977) advocate a gradual

move from linguistic form to communicative competence as they saw it. They were among the first to design functional syllabuses in the US.

Communicative approaches can be broadly defined as a philosophy of language and its corresponding approach to language education, in which language is seen as communication. Raimes (1983) analysed language teaching materials in the US when English functional syllabuses were first becoming prominent. She sees a conflict of the old (audio-lingualism) and the new (communicative language teaching) paradigms. This meant that many teaching materials were still firmly rooted in audiolingualism, with its focus on grammar and syllabus-controlled learning of vocabulary. Raimes concludes that teaching materials may have changed in design and procedure but not in the sense of a changed underlying philosophical approach. She mentions the 'overlay solution', the 'label change solution', and the 'add-a-component solution'. To illustrate these labels, we will now discuss how Paulston, Rivers and Valette interpreted communicative ability.

According to Berns (1990), Paulston's interpretation of communicative competence is an instance of what Raimes had called "an overlay solution". It meant that the contents of a textbook were divided into grammatical structures, to which a list of functions were added. Paulston's alternative to previous syllabuses are communicative drills of the following kind:

> *Teacher: Describe the weather in your country.*
> *Student: It's (beautiful/wonderful)*
> *Teacher:What is your responsibility?*
> *Student:My responsibility is (to learn English/learning English)*
> (Berns, 1990: 84/5)

The main change with the usual drills is that they require the learners to answer truthfully. Paulston did not see communicative competence as a necessary goal of a language programme. Yet, she took up Hymes's elaboration of the social rules of language use "rather than taking it to mean simply linguistic interaction in the target language." (1974: 374). She felt that communicative practice of linguistic forms in the classroom was often devoid of any social meaning. Moreover, she was of opinion that the manipulation of linguistic forms did not accurately reflect L2 culture. The poignant question here, also demonstrated by the sample dialogue above, of course is: whose L2 culture? To start with, there are considerable differences between British English cultures or American English cultures and other " Englishes" around the world. Besides, English is increasingly used as a lingua franca in a variety of intercultural encounters. Savignon rightly points out that also the language classroom itself has a culture of its own that should be taken into account.

Rivers (1972) put aside considerations of cultural specificity and focused on *spontaneity of expression*. Rivers was in favour of a progression from controlled structure practice to creative use of language for communication. She considered language acquisition as going from production exercises, drills, or activities, whereby the learner concentrates on formal accuracy, to interaction for a communicative purpose. Rivers felt it was up to the teacher or the teaching materials to structure this interaction in line with the grammar and idioms that had already been presented. She believed that " For any aspect of language structure a game or simulated activity can be invented which forces the learners into autonomous activity in which they produce the same type of responses as in an artificial teacher-directed exercise." (1972: 31).

Berns and Raimes relativise Rivers's use of "autonomous interaction". They see her as a representative of a "label-change solution". Rivers came to terms with the new concept of communicative competence by relabelling the terminology, but not the audiolingual concepts. She saw language learning as a moving from skill-getting to skill-using. Yet, the skill-using stage was still based on a paradigm that was structurally linguistic in its philosophy and typically audiolingual in its methodology.

Spontaneous language use was still controlled and aimed at linguistic perfection. In later models (1983), Rivers would add Halliday's functions as an added component to her basically audiolingual curriculum.

A final representative of a *from surface structure to meaning* interpretation of communicative competence is Valette (1977), who focused on L2 programme goals. She presented a sequential list of five goals that organized objectives from the simplest of behaviors (Stage 1) to the most complex (Stage 5). Her list echoes audiolingual methodology as well:

1. ***Mechanical skills****: The student performs via rote memory rather then by understanding.*
2. ***Knowledge****: The student demonstrates knowledge of facts, rules, and data related to foreign language learning.*
3. ***Transfer****: The student uses his knowledge in new situations.*
4. ***Communication****: The student uses the foreign language and culture as natural vehicles for communication.*
5. ***Criticism****: The Student analyses or evaluates the foreign language or carries out original research.*
   (1977: 19-20)

Advocates of the first general perspective distinguish between skill-getting and skill-using, and see skill-getting as the first and necessary step in language learning. What they do not concentrate on is a recognition of skills such as the expression, interpretation and negotiation of meaning. This brings us to a second interpretation of communicative competence.

4.4.2 From Meaning to Surface Structure

A number of scholars and practitioners embraced the new concept of communicative competence in more radical ways. We will subsequently refer to Widdowson, Piepho and Savignon and see how they interpreted communicative competence.

Widdowson's discourse-based approach

Savignon (1983,1997) and Berns (1990) mention Widdowson and his discourse-based approach as a representative of a perspective on communicative language education that considers meaningful communication and the interpretation of meaning as the starting point. Widdowson feels that actual use of the language will most effectively trigger the learner to study surface structure or usage rules.

For Widdowson (1978) the first step towards acquiring communicative competence is the interpretation of discourse. Widdowson felt that " The learner's task (is) one which involves acquiring a communicative competence in the language, that is to say, an ability to interpret discourse whether the emphasis is on productive or receptive behaviour." (1978: 144). He carefully distinguishes between *linguistic skills* and *communicative abilities.* Widdowson is of opinion that even though the acknowledged aim of language instruction is communicative competence, the traditional focus has been on linguistic skills. This does not in itself ensure the development of communicative abilities. " On the contrary, it would seem to be the case that an overemphasis on drills and exercises for the production and reception of sentences tends to inhibit the development of communicative abilities." (1978: 67).

Widdowson's interpretation of discourse directly relates to the needs and present knowledge of the learner. He sees L2 acquisition as most efficient when learners rely on their existing knowledge of the world, i.e. on previously acquired aspects of communicative competence, to interpret meaning within a new linguistic code. He felt that linguistic skills and communicative abilities should never be treated in isolation. Just as one learns to sail by sailing or to paint by painting, one learns to

communicate by communicating. Widdowson put this as “What the learner needs to know how to do is to compose in the act of writing, comprehend in the act of reading, and to learn techniques of reading by writing, and techniques of writing by reading.” (1978: 144). Widdowson aims to have learners develop competence in the *use* of the rule system of the language. Analysing discourse involves the learners in searches for meaning and interpretations of texts. In order to understand texts learners have to organize their means of analysis. Language *use* will provide the learner with incentives to concentrate on linguistic skills, such as language *usage*. That way, another language is acquired most effectively.

Piepho’s “Kommunikative Didaktik”

Another interesting representative of a *meaning to surface structure* approach to communicative language teaching is discussed by Berns (1990). She mentions Piepho’s communication-based approach as an orientation that also focuses on meaningful communication. Piepho’s approach is different from Widdowson’s. German methodologist Piepho (1974,1979) was largely responsible for new orientations in the teaching of German as a first and second language as well as of English as a foreign language. We will briefly discuss his approach in the next paragraphs.

Piepho’s approach to English language teaching is characterized by two main educational concerns. His first concern is to reduce to its fullest possible extent the discrepancy between school and the reality of socialization. Opportunities for language socialization should be created in the English lessons within the German context. Piepho’s second educational concern “is establishing English language teaching and learning as a means of developing a careful and systematically sequenced approach toward expression, interpretation, and negotiation within learners in their experiences with the sociocultural reality of English.” (Berns 1990: 96). Thus, Piepho links meaningful communication directly to social and cultural aspects of English, which are experienced and subsequently interpreted, negotiated, and expressed. Piepho introduced his view of communicative language didactics or “Kommunikative Didaktik’ in the 1970s. In those years they were not generally accepted and proved to be controversial. Berns feels that this was “in large part due to his use of Jürgen Habermas’s (1970, 1971) social theory and interpretation of communicative competence as a basis for his communicative approach.”(1990: 96). Piepho did not primarily aim at developing communicative language teaching materials. “His objective was to democratise language teaching and to break down elitist barriers to the development of communicative competence for all learners.” (1990: 96). Piepho’s objective reflects Habermas’s socio-political view of communicative competence.

Habermas is concerned with the notion of communicative competence in ideal terms. He sees communicative competence as the mastery of an ideal speech situation. Habermas considers a communicative act ideal, when the actual motives of the hearer are identical with the linguistically apprehensible intentions of the speaker. Its prerequisite is an unhindered agreement between the participants about the thematic and situational parameters, the inherent meaning relationships, and the social conventions appropriate to the context (Berns, 1990: 97). The statements in an ideal dialogue are assumed to be true, appropriate, sincere, and comprehensible. The ability to enter in such a dialogue is labelled as communicative competence. Ideal communication fails if one of the participants questions the truth, appropriateness, sincerity or comprehensibility of (parts of) the dialogue. “At this point, discourse begins, which is the discussion (negotiation) between the participants oriented toward reestablishment of agreement on basic principles, reaching a consensus, or resumption of communication. This distinction between pure communication and discourse is central to Habermas’s theory.” (Berns, 1990: 98). Habermas qualifies speech situations in two ways: (1) the potential to produce

an ideal speech situation and (2) knowledge and competence in role behaviour, which he calls “symbolic interaction” (1971).

Habermas’s view of ideal communication resounds in Piepho’s definition of communicative competence. He feels communicative competence is :

> *the ability to make oneself understood, without hesitation and inhibition, by linguistic means which the individual comprehends and has learned to assess in terms of their effects, and the ability to comprehend communicative intentions even when they are expressed in a code which the speaker himor herself does not know well enough to use and is only partially available in his or her idiolect.*
> (1974: 9-10, Translated from German by Berns)

“Kommunikative Didaktik” simultaneously promotes learning about a subject and about the processes of social interaction (Edelhoff, 1983). In addition, the development of critical thinking is inherent in Piepho’s approach. Piepho and fellow progressive educators aim at:

> *the creation of a more democratic base in the schools and the breakdown of elitist barriers to a quality education which would recognize, among other things, the legitimacy of language variety and deviation from the norm as well as the necessity of developing critical thinking in the learners. This critical thought would be expressed through language, whether it be in the first or second language classroom. The relationship between these aims is realized when learners are encouraged to engage in discourse in Habermas’s sense, that is, challenge, criticize, and suggest improvements in the status quo, be it of class texts, activities, or the society at large, with the language being learned as the medium for the discourse.*
> (Berns, 1990:99)

Both Widdowson and Piepho can be called representatives of a meaning-to-surface-structure approach. Advocates of such an approach believe that grammatical and formal exercises should be generated by the needs that stem from communicative experiences. Language practice and discrete exercises are most useful when they accompany or follow rather than precede communicative experiences. Proponents of this general interpretation of communicative competence do not believe in an atomistic or sequential view of language learning. A final example of a from-meaning-to-surface-structure approach is a language pedagogy suggested by Savignon.

Savignon’s interactional approach

Savignon’s approach (1983, 1997) is based on the view that “meaning making” and language are inextricably bound. Mastery of a language involves knowing how to use it for creating meaning as well as knowing about its forms. Inspired by Jakobovits and Hymes and underpinned by the theoretical models developed by Canale and Swain (1980) and Canale (1983), Savignon extended and reinterpreted Hymes’s concept of communicative competence. She presented communicative competence as a viable pedagogical construct in second and foreign language education. As we have quoted before, Savignon sees communication as ‘a continuous process of expression, interpretation, and negotiation of meaning’ (1997:17). She stresses the interactional and sociocultural nature of the competences involved in human communication and suggests a communicative curriculum in which language is taught for communication. The curriculum she proposes has five potential components that lack hierarchy and sequence and are interdependent. Below we will briefly present the five components.

- ***Language Arts***: analysis of formal language characteristics and their use.

  Language Arts refers to what language teachers have generally learned to do best, i.e. the analysis and applications of formal language characteristics, such as grammar, spelling, and vocabulary. Savignon has included the term arts because she feels language learning is more related to arts than to science. In this component the focus is on usage rules and explanations how language works. It involves systematic practice e.g. by way of spelling tests, vocabulary expansion exercises, pronunciation exercises and a variety of language games. The component should be seen in relation to the other four and is no more or less important than the others.

- ***Language for a purpose****:* meaningful use of language to achieve real and immediate communicative goals.

  The second component relates to the creation of opportunities to use the target language in meaningful and goal-related ways. These goals will vary from individual to individual. Using the target language as the classroom language is one way of introducing learners to meaningful communication.

- ***My language is me: Personal L2 use***:
  Psychologically and intellectually engaging the learners in self-expression.

  Learners are encouraged to express feelings, thoughts, attitudes, beliefs etc. regarding the target language and the cultures in which the language is used. Intercultural awareness is stimulated by having the learners write personal journals, supply cultural information about their own lives and study target language culture. Thus, personal L2-use is meant to enhance language learning motivation.

- ***"You Be, I'll Be": Theater Arts*****:**
  "All the world is a stage", and by extension so is the classroom: play and drama as curricular components.

  The fourth component includes a variety of drama activities, such as role play, simulations, and rehearsal and gives the learners a platform for target language practice and use. In addition, it creates the opportunity to analyse language.

- ***Beyond the Classroom:***
  Exploration of the L2 community and/or bringing the L2 community in the classroom.

  This final component invites the learners to explore L2-communities outside their classroom environment. It comprises the exploration of either real or virtual L2-settings.

Savignon advocates a language pedagogy that attempts to involve the learners to the full in language use. Just like Widdowson and Piepho, Savignon focuses on the language *learning process* and a corresponding language pedagogy. Their

examples bring us to a final and typically European interpretation of communicative competence, which focuses on, often detailed, specification of context.

4.4.3 Specification of Context

The final general perspective on communicative language competence discussed by Savignon is related to the *selection* of the language to which the learner is exposed in particular social contexts. Specification of social contexts involves the creation of a taxonomy of knowledge and skills, functions and notions, curricular objectives and descriptors of communicative competence. The specifications are not accompanied by communicative teaching or assessment strategies. How another language is learned is not its primary concern. The need to specify communicative context arose from the necessity to provide a pragmatic alternative to the grammatical syllabuses that had persistently dominated foreign language education.

Savignon points out that lots of language textbooks are in fact grammar books. Formal or structural criteria are often embedded in texts in such artificial ways that the only conceivable context of a text is the textbook itself. An example of a grammar demonstration dialogue is from Rivers (1975: 22):

Paule:	*Où vas-tu ce soir? [Where are you going tonight ?]*
Madeleine :	*Je vais en ville avec ma famille. Nous allons au cinéma. [I'm going downtown with my family. We're going to the movies.]*
Paule:	*Qu'est-ce que vous allez voir? [What are you going to see ?]*
Madeleine:	*Zazie dans le Métro. Mes cousins vont voir le même film demain. [My cousins are going to see the same film tomorrow.]*

The dialogue has been quoted as a poor example of authentic speech transaction. Details of the context are missing. Is Paule looking for something to do? If so, Paule would have asked Madeleine something like "*Qu'est-ce que tu fais, toi, ce soir*?" [What are you doing tonight?]. Or if Paule responds to an earlier statement by Madeleine, who says she is going out tonight, Paule's likely query would have been "*Ou Ça*?" [Where?]. Valdman (in Savignon, 1997:38) had wryly noted "Surely the purpose of her question is not to provide *Madeleine* with the opportunity to rehearse the present indicative of *aller*." Critics such as Valdman called for the development of teaching materials that provided viable alternatives to curricula that were dominated by grammar and isolated idioms learned and reproduced without any communicative context.

Social contexts defined

Seminal work in defining social contexts of foreign language communication and subsequently developing assessment scales and descriptors has been done by the Council of Europe.

From the outset the Council had opted for plurilingualism in a unifying Europe. It was recognized from the start that the languages spoken in Europe constituted an essential part of the European Cultural Heritage. One of the issues that had to be faced was the inability to understand one another in a unified Europe. For this reason the Council of Europe has long seen the promotion of foreign language education as a priority area, with the development of intercultural awareness seen as an essential part of the development of knowledge and skills in another language or other languages. (Morrow, 2004).

In the late 1950s one of the first concrete measures taken by the Council was their support for the development of *Le Français Fondamental*, a specification of a basic grammar and vocabulary for French, and *Voix et Images de France*, an audiovisual course of French for adults.

Initially a lot of work was done by the Council's *Committee of Out of school Education*. Their work led to the development of notional-functional syllabuses, the Threshold Level (Van Ek & Alexander, 1975), and the birth of approaches that were more communicative than the previous grammatical syllabuses had been. Guidelines for the development of language teaching materials were developed that were supposed to offer viable alternatives to the grammatical syllabus (See Trim, 2001). Such guidelines for syllabus design were based on notions associated with various language functions. Quoting Wilkins, Savignon (1997:13) explains what notional-functional syllabi comprise:

> *The term notional means semantic and is borrowed from linguistics, where I denotes grammars based on semantic, or meaning, criteria rather than on structural, or formal, criteria. Wilkins defines a notional syllabus as "any strategy of language teaching [emphasis added] that derives the context of learning from an initial analysis of the learner's need to express such meanings." (1976, p.23). He refers to this approach to the specification of the language to be taught as a communicative approach. Communicative function or social purpose, determines the notional, or semantic, features of an utterance. The Council of Europe definition of a "threshold level" of language proficiency for adults with identified L2 needs specifies the following components:*
>
> 1. *the situation in which the L2 will be used, including the topics that will be dealt with;*
> 2. *the language activities in which the learner will engage;*
> 3. *the language functions that the learner will fulfil;*
> 4. *what the learner will be able to do with respect to each topic;*
> 5. *the general notions that the learner will be able to handle;*
> 6. *the specific (topic-related) notions that the learner will be able to handle;*
> 7. *the language forms that the learner will be able to use;*
> 8. *the degree of skill with which the learner will be able to perform*
>
> [Van Ek, 1975:5]

A *unit-credit* system was developed that would enable learners to study 'units' of work in specific notional-functional areas and gain 'credits' for these. This is where Savignon's (1983,1997) example of a 'specification of context' interpretation of communicative language ability stops. However, the works and ideas of pragmalinguists such as Van Ek and his colleagues did not. Through the 1980s the Council of Europe was involved in a number of initiatives that built on the ideas incorporated in the Threshold Level. Two examples are *The Waystage*, a lower-level set of objectives, and the design of a multimedia TV course called *Follow Me.* From 1989-97 a major project was undertaken called *Language Learning for European Citizenship.* It aimed to provide guidelines for the reform and development of language teaching in the member states of a growing Europe. The project resulted in the publication of *A Common European Framework* (CEF) *of Reference for Languages: learning, teaching and assessment*, which was first published as a report in 1996, revised in 1998 and published in a commercial version in 2001 (Council of Europe, 1998,2001). The CEF was developed under the auspices of Trim, Coste and North. Because we feel the CEF is both an influential and a comprehensive example of a specification of the context of language use, we will discuss the framework in more detail. We will first refer to its formal origins, comprehensiveness, intended goals, objectives and much debated practicality and implementation (see also Morrow, 2004; Fulcher, 2004, North, 2004, Davidson & North, 2006). In a final

paragraph, we will deal with the foreign language user's competences as they have been specified in the CEF.

The Common European Framework of Reference

The formal origins of the CEF go back to 1991, when it was agreed that:

> *the mutual recognitions of qualifications, and communication concerning objectives and achievement standards would be greatly facilitated if they were calibrated according to agreed common reference standards, purely descriptive in nature.* (Trim, 2001:5)

The CEF developed into a comprehensive and weighty document with a far wider application than its original aim by its codification and examination of what Trim calls:

> *language use and the many 'competences' i.e. the shared knowledge and skills which enable users of a language to communicate with each other. wherever possible, these are separately calibrated with brief descriptors defining six levels of proficiency.* (Trim, 2001:5)

Morrow (2004) refers to its comprehensiveness both as a strength and a weakness. The CEF attempts to document the many competences deployed in communication, and defines different levels of knowledge and performance in these competences. For one thing, the Framework makes explicit the complexity of foreign language education and communicative competence as a construct. The CEF is not without its critics, as Morrow has shown. Course or test designers, materials writers, teacher trainers, teachers and learners are overwhelmed and often find:

> *the sheer amount of detail, the range of descriptors, and the plethora of terminology completely baffling. And it has to be said that the published versions of the CEF are not exactly user-friendly. There is little to guide the first-time reader around the material; the print is small, the layout dense and 'heavy', the language itself is ponderous and often convoluted; specialist terminology abounds, and is often used in ways which seem idiosyncratic- and there are seemingly endless tables and descriptors whose relationship to one another is very difficult to discern.* (Morrow, 2004:7)

Despite these drawbacks, the CEF lives up to its intended goals. It is important to read the document and reflect on its comprehensive subject matter in the ways envisaged by the authors. This is how Trim et al. present the goals of the publication:

> *You may, of course, use the Framework document in any way you wish, as with any other publication. Indeed, we hope that some readers may be stimulated to use it in ways we have not foreseen. However, it has been written with two aims in mind:*
>
> *1. to encourage practitioners of all kinds in the language field, including language learners themselves, to reflect on such questions as:*
> - *what do we actually do when we speak (or write) to each other?*
> - *what enables us to act in this way?*
> - *how much of this do we need to learn when we try to use a new language?*
> - *how do we set our objectives and mark our progress along the path from total ignorance to effective mastery?*
> - *how does language learning take place?*
> - *what can we do to help people learn a language better?*

2. *to make it easier for practitioners to tell each other and their clientele what they wish to help learners to achieve, and how they attempt to do so.*

> *One thing should be made clear right away. We have NOT set out to tell practitioners what to do, or how to do it. We are raising questions, not answering them. It is not the function of the Common European Framework to lay down the objectives that users should pursue or the methods they should employ. This does not mean that the Council of Europe is indifferent to these issues. Indeed, a great deal of thought and work has been put into the principles and practice of language learning, teaching and assessment over the years by colleagues in our member countries working together in the Council of Europe's Modern Languages Projects.*
> (Council of Europe, 1998:V)

The CEF aims at having all foreign language practitioners reflect on issues that matter in the field of foreign language education with the intention to become more specific about *what* is learned, *how* this is done and *why* this is done in particular ways. As we will see, its approach is far from unprincipled. The document has tried to do justice to both older and more recent models and theories of learning, teaching and assessment. Nevertheless some critics, such as Bausch, Christ and Königs (2002), have argued that the CEF is weak on theoretical grounds. Morrow (2004) cites an anonymous reviewer at the German Amazon website of Bausch, Christ and Königs' book who says:

> *Es ist mein Eindruck, dass nicht alle den Referenzrahmen gut gelesen haben. [I have the impression that not all of the contributors to this book have read the CEF closely.]*

Morrow had an anonymous reviewer speak for many. Even though we welcome a critical attitude towards the CEF, we feel that critics do not always do justice to the thirty-odd years it took for the framework to develop and to its enormous potential.

Whatever one's view of the theoretical underpinnings of the CEF is, what remains is an impressive effort to help practitioners reflect on matters that will make FL teaching, learning and assessment both more specific and more effective in the sense that a great many opportunities to improve the quality of teaching and learning are created.

Competences specified in the CEF

We will now turn to how a learner's communicative competence has been defined in the CEF. The user/learner's competences are elaborately presented in chapter 4 *Language Use and the Language User/Learner* (Council of Europe,1998). In that chapter, the notion of communicative competence has been embedded in a range of general competences. such an interpretation of communicative competence can be better understood if we take one of the objectives of the CEF into closer consideration. It states that:

> *To promote methods of modern language teaching which will strengthen independence of thought, judgement and action, combined with social skills and responsibility.*
> (Council of Europe, 2001:4)

Independent thought, judgment and action and *social skills and responsibility* represent values that fit in both with a model of learner autonomy as a pedagogical construct and models of communicative competence that focus on meaningful interaction. Widdowson's discourse-based approach, Piepho's socio-politically oriented communication–based approach, and Savignon's interactional approach, were the examples we discussed. Savignon and Piepho in particular value

responsible self-expression as an integral part of how they view communicative competence.

According to the CEF (Council of Europe, 1998: 39), foreign language users and learners draw upon a number of competences developed in previous experience when they carry out certain tasks and activities required to deal with the communicative situations in which they are involved. Participation in those events, including events that promote language learning, results in the further development of the learner's competences for both immediate and long-term use. The CEF acknowledges that all human competences in one way or another contribute to the ability to communicate and may be regarded as aspects of communicative competence. However, the authors consider it useful to distinguish the more *general competences* from *communicative language competences*. In addition, they see linguistic competences as only one of the competences a learner needs to be able to communicate successfully in a foreign language.

The authors of the CEF specify a number of general competences that are not linked to the formal systems of grammar and idioms of the foreign (or second) language someone intends to learn and use. The authors do have a point. Learners who start learning and/or using a second or a foreign language do not start from scratch. They already have the experience of learning and using at least one other language. From infancy they have learned to communicate with their environments using symbols, the native language in particular.

The general competences that the authors of the CEF mention are the knowledge and abilities all human beings generally acquire or learn in one way or another. The CEF distinguish four types of general competences: *declarative knowledge*, *skills and know-how*, *existential competence*, and the *ability to learn*. Together these general competences represent what a learner already knows and is able to do in varying degrees when (s)he starts learning another language.

It is not just the general competences that receive attention in the CEF. The definition and specification of the general competences is followed by an even more comprehensive specification of the so-called *communicative language competences*. These competences are directly linked to the formal systems of the language to be learned and used.

The CEF can be seen as a document that specifies the contexts offered by e.g. Canale & Swain's and Bachman's models of communicative competence in a detailed way. The CEF (1998) divides communicative language competences into:

- **linguistic competences**
  These competences comprise specifiable *lexical*, *grammatical*, *semantic* and *phonological* competences.

- **sociolinguistic competence**
  Specified are *markers of social relations*, *politeness conventions*, *expressions of folk-wisdom*, *register differences* and *dialect and accent*.

- **pragmatic competences**
  They comprise *discourse*, *functional* and *schematic design* competences.

In line with Bachman's (1990) model of communicative competence, strategies are seen as a separate category. The CEF (1998:58-61) identifies production, reception, interaction and mediation strategies.

Our discussion of the CEF closes off our section on interpretations of communicative competence labelled as *specification of context.* The three interpretations of communicative competence we discussed focused on *form*,

meaning or *context*. These ways of understanding might suggest a lack of consensus in how communicative competence is to be interpreted in communicative language education. Yet, we would like to argue that the three interpretations are not necessarily mutually exclusive. A focus on form may well be an integral part of a focus on meaning, and specifications of the social context may help to unravel and better understand what foreign language education is about. Because we would like to provide consensus rather than controversy regarding the implementation of communicative competence in foreign language education, we will now refer to aspects of communicative competence in foreign language education on which there is general agreement.

4.5 Communicative competence in foreign language education

The implementation of a complex construct such as communicative competence in foreign language education is a formidable task. It seems very much dependent on the assumptions and beliefs that researchers and educators have about effective foreign language education. We feel it is useful to call attention to the assumptions and beliefs related to communicative ability in foreign language learning and teaching on which there is general agreement. Such common denominators may help to specify what communicative language education, with its central construct of communicative competence, involves. Berns (1990) mentions eight aspects.

1. *Language teaching is based on a view of language as communication. That is, language is seen as a social tool which speakers use to make meaning; speakers communicate about something to someone for some purpose, either orally or in writing.*
2. *Diversity is recognized and accepted as part of language development and use in second language learners and users as it is with first language users.*
3. *A learner's competence is considered in relative, not absolute, terms of correctness.*
4. *More than one variety of language is recognized as a model for learning and teaching.*
5. *Culture is seen to play an instrumental role in shaping speakers' communicative competence, both in their first and subsequent languages.*
6. *No single methodology or fixed set of techniques is prescribed.*
7. *Language use is recognized as serving the ideational, interpersonal, and the textual functions and is related to the development of the learners' competence in each.*
8. *It is essential that learners engage in doing things with language, that is, that they use language for a variety of purposes in all phases of learning. Learner expectations and attitudes have increasingly become recognized for their role in advancing or impeding curricular change*
   (Berns, 1990: 104).

What we see here are assumptions, beliefs and recognitions that are as relevant for discussion today as they were in 1990. A basic premise in communicative language education seems to be the view of 'language as communication', in which language is seen as 'a social tool to make meaning'. Such a belief implies a number of recognitions. It is recognised that communication has content as well as purpose and that people communicate orally as well as in writing. In addition, it is recognised that language development and use are diverse, involving more language varieties

than the single one offered by an alleged native-speaker model. Finally, it is acknowledged that language use serves language functions and is related to the learner's development and competence in each. In view of the acceptance of diversity, one cannot but consider a learner's competence, which has been defined as multidimensional and multifaceted, in relative terms. Related to the acceptance of diversity are the belief that more than one methodology or set of techniques can be used to achieve communicative ends and that culture shapes and frames a person's communicative ability. A final and crucial view Berns mentions is the need to actively engage learners in using the language in a variety of ways.

The assumptions, beliefs and recognitions mentioned above have given way to a host of studies of communicative ability in education. We will end our second theoretical chapter with some important research insights and foci related to communicative competence and CLE. The overview has largely been based on Musumeci (2004).

4.6 Research insights into communicative language education

Research on communicative language education has been seeking answers to a number of poignant questions. (Musumeci, 2004:90). What is the relationship between language learning and communicative language teaching? How is it possible that students can spend years in language classes and never attain enough proficiency to meet their communicative needs? What has over thirty years of research in second language acquisition (SLA) contributed to our knowledge what teaching towards communicative competence looks like? It is important to state that insights into the mechanisms of SLA generally support principles of communicative language education. There is indeed empirical evidence that supports tenets of CLT, which views learners' ability to interpret, express and negotiate meaning in another language as both the process and product of instruction (Musumeci, 2004:89). We will highlight some important studies and their findings in three sections entitled *a focus on meaning*, *a focus on form*, and finally *a focus on cultural aspects of language use*.

4.6.1 A focus on meaning

The successes of Canadian-French immersion programmes provided convincing empirical evidence for the positive effects of engaging classroom learners in meaningful L2 communication on their communicative competence (Harley & Swain, 1984, Lapkin, 1984; Swain, 1985). The researchers reported that students who had been taught subject matter in and through the medium of French, by teachers who were willing and able to make such language accessible and understandable to the learners, outperformed on every measure the students who were taught French in the traditional way. This seminal example of second language acquisition research in the 1980s was characterized by a cognitive and psycholinguistic focus on what was going on into learners minds when they were trying to make sense of the language data into which they had been immersed. Due to the influence of Krashen (1982, 1985), research very much concentrated on the examination of 'comprehensible input' and what teachers can do to ensure that input becomes 'intake', i.e. what the learners extracts from the input to create a personal grammar. Research very much concentrated on the teacher/native speaker and his/her talk, with the learner seen as the novice recipient (Long, 1983, Wesche & Ready, 1985; Larsen-Freeman & Long, 1991). Yet, despite the relative success of the immersion programmes, it was shown that comprehensible input was a necessary, but in itself an insufficient condition for language acquisition (Long, 1990). These findings stimulated further research into the possible sources of learners' gaps in performance, as researchers attempted to discern whether they were attributable to deficiencies in the input (Swain, 1988 &

1993), or a failure to provide a focus on form in communicative classrooms (Lightbrown & Spada, 1990).

With regard to error correction and feedback, researchers (Ellis, 1985, 1997) made a discovery that has tremendous consequences for traditional language teaching uncritically focusing on grammar. Explicit grammatical instruction, including error correction, appears to have very limited effects. Instead, grammatical structures appear to be acquired according to an immutable developmental sequence, the route of which is unaffected by instruction. These results fall in with a concern expressed by many reflective teachers. Even though structures have been taught and tested with some success, learners seem to 'forget' them once the focus shifts from explicit attention to spontaneous language use. However, more recent reviews of research on form-focused instruction (FFI) report that it can contribute to the acquisition of implicit knowledge. Two variables appear to influence its success: the choice of the target structure and the extent of the instruction. FFI involving extensive instruction directed at "simple" structures is more likely to succeed. Yet, limited instruction directed at complex structures can also be successful, provided that the target structures are readily availaible in non-instructional input (R.Ellis, 2002).

Swain's essentially form-based output hypothesis (1985) drew attention to the relationship between form and meaning in language learning. In discussing the results of Canadian French immersion education and their consequences for SLA, Swain (1985) put forward that in addition to comprehensible input, successful language acquisition required opportunities for 'comprehensible output' on the part of the learners as well. Such pushed performance would force learners to refine their output to make it more closely match native-like models (Musumeci, 2004:92). Accordingly, output lets learners "test hypotheses about L2, experiment with new structures and forms, and expand and exploit their language resources" (Pica et al., 1989:64). Although the concept of output is still firmly rooted in a psycholinguistic framework, the notion of negotiated interaction is highlighted as well. As compared to earlier SLA-research, the focus of attention was shifted from what the teacher/native speaker does with the language to what the learners do with it. Negotiated interaction is generally seen as an essential factor in the language learning process. In order to learn how to use a language, one actually has to use the language in meaningful ways. Meaningful interaction in the target language provides learners with opportunities to test their hypotheses about how language works. Kowal & Swain (1997) have posited that negotiated interaction may also serve to move learners from a system of purely semantic processing into one that demands syntactic processing of the input. Musumeci (1996) stresses that the amount of negotiation that actually takes place in ordinary classroom interaction is an area for continued investigation.

Swain's findings corroborate earlier studies carried out by researchers working within a sociolinguistic framework, who, in line with Hymes, primarily focused on language as social behaviour. They looked at learners' use of language as an essential factor in the acquisition process, concentrating on the discourse level rather than on the word or sentence level (Hatch, 1978, 1992). In a landmark study, Savignon (1972) had shown that university students learning French as a foreign language who spent just one of five class sessions per week engaged in meaningful, communicative acts outperformed students in conventional classes on a test of communicative ability. These results contradicted the logic at the time, i.e. that the study and practice of grammar leads to functional language ability. Savignon's study highlights the importance of including communicative tasks in the curriculum if communicative competence is the desired outcome. The nature of such tasks have been and continue to be vital areas of research, whether the tasks are teacher-fronted or learner-centred, conducted in a whole-class setting, in pairs, or in small groups, whether they require one-way or two-way communication and how characteristics of individual learners may affect their performance during the tasks (Gass & Varonis, 1985; Crookes, 1989; Pica et al., 1989; Young, 1999; Ortega, 1999).

Lee (1999) describes task-based instruction based on research findings. The research on the nature of communicative tasks helps to provide guidelines for the creation of tasks that have the learners communicate and learn in meaningful ways. In addition, the studies reveal how the complex interaction of task attributes, the roles and characteristics of the participants, and the organizational structure of the task itself may affect the outcome of the task, and presumably, of learning (Zuengler, 1989; Zuengler & Bent, 1991; Foster, 1998; Dörnyei, 2002).

4.6.2 A focus on form

Some of the studies referred to above have elucidated the limited role that explicit instruction in grammar plays for the acquisition of structural features. Other studies have shown that form-focused instruction can be effective. The precise nature of the contribution that explicit instruction makes in acquiring another language, continues to be a matter of debate. Some researchers have argued that even though grammatical structures appear to be acquired in an immutable developmental sequence, the rate of that development can be positively influenced by instruction. Current research that focuses on form investigates what types of intervention may be effective and when (Mackey & Phlip, 1998; Spada & Lightbrown, 1999; McCollam Wiebe, 2002).

Yet, the current interest in focus on form tends to address only the learner's linguistic performance. Little attention is paid to language as social behaviour, particularly with regard to the appropriateness of utterances. In addition, recent research on form continues to rely heavily on a native-speaker norm to measure learners' success. Identifying who is a native speaker (Wong-Filmore, 1992) and what exactly constitutes a target community (Hartford, 1997) proves to be problematic. In that light, Berns (1990) reminds us of the validity of maintaining diversity in both models for and measures of learners' performance in the communicative curriculum. An interesting and informed change in approaching grammar in communicative language education is suggested by Larsen-Freeman (2003). She coined the term *grammaring*, which refers to grammatical instruction as a "process of doing grammar". This process is meant to result in learners using grammar structures accurately, meaningfully, and appropriately. Larsen-Freeman considers grammaring as "the fifth skill" (alongside listening, speaking, reading, and writing). Grammar lessons are no longer about knowing language systems (declarative knowledge), but about knowing how to use language (procedural knowledge). Compared to those who wish to dismiss a focus on form altogether, Larsen-Freeman takes a fairly intermediate position by stating that there needs to be both explicit and implicit instruction, and inductive, deductive, and abductive (after the fact inferencing of why something is true) learning as well. However, in keeping with the communicative approach, she does not advocate too much teaching of grammar. Instead, she puts forward that by helping students learn how to figure out their language choices in grammar, they will be empowered to better use the language. Students learn to figure out their choices by passing through three alleged dimensions of form, meaning/semantics, and pragmatics/use. Larsen-Freeman's interest in pragmatics, links up with studies in the field that focused on explicit instruction (Billmyer, 1990; Bardovi-Harlig & Hartford, 1993). Markee (2000) advocates the necessity of including conversational analysis as a method of SLA-research in view of insights it is likely to provide in the language learning process, especially with regard to how learners acquire meaning in conversational contexts. Markee favours experimental, largely quantitative, studies to that effect.

4.6.3 A focus on the cultural aspects of language use

Several researchers (McCarthy & Carter, 1994; Van Lier, 1995; Schäffner & Wenden, 1995) have argued that the educational potential of language study should not be limited to units of pragmatics language use ability. In addition, learners should

be encouraged to reflect on the cultural aspects of language use, particularly its power to highlight certain interpretations of reality or the social organizations of its forms. This way language teaching becomes “useful” to the degree that it helps people to make the judgments needed not only for engaging in verbal exchanges with others, but also for understanding them (Kinginger, 2004: 102). Kramsch (1993) is among the scholars who have expanded on notions like these. She has described an experiential approach to learning language as culture. Kramsch and many with her (e.g Byram, 1997, Byram & Zarate, 1997, Sercu, 2000, 2002; Elsen & St.John, 2007) feel that in addition to facts about speech communities, learners need access to discovery processes that help them to contrast and compare not only the target language cultures, but also the very culture that has shaped their feelings and thoughts. It helps learners to cultivate an understanding of their own positions as observers of other ways of life. Kramsch feels such an experiential approach is characterised by dialogue.

> *Through dialogue and the search for each other’s understanding, each person tries to see the world through the other’s eyes without losing sight of him or herself. The goal is not a balance of opposites, or a moderate pluralism of opinions, but a paradoxical, irreducible confrontation that may change one in the process.*
> (Kramsch, 1993:231)

An interest in dialogue and interaction is also shared by researchers working within a sociocultural framework. The revival of sociocultural theory has led to an ongoing development of Vygotsky-inspired research and its application to second- and foreign-language developmental processes and pedagogies. Students particularly focus on the interaction between novice and expert as the initial stage of learning, a stage that prefigures restructuring of the individual’s cognitive system (Aljaafreh & Lantolf, 1994, Lantolf, 2004, Grabois, 2004, Thorne, 2004, Lantolf & Thorne, 2006). Concepts are embraced such as ‘collaborative dialogue’ and ‘languaging’ (Swain, 2006). According to Swain these concepts are still related to output, but the notions stress the co-construction of language as a process (languaging), instead of as a product (language). “The kinds of activities that we’ve been researching engage the students in speaking and writing and reflecting on what they have said or written – that is, co-constructing meaning through languaging” (2006:2). The verb *languaging* refers both to activity and process. As an activity, languaging is an integral aspect of the way humans think and their meaning-making selves. Languaging is what people do to transform their thoughts into a shareable resource – shareable with themselves and others. Swain argues that languaging is part of the second language learning process. Agency may facilitate or hinder the process of second language learning. Languaging and agency are made visible in collaborative settings, but they can also be observed in settings where individuals are alone. Languaging and agency are also concepts that have sensitised scholars such as Van Lier (2007), with his more recent focus on action-based teaching, autonomy, and identity.

4.7 Summary

In this chapter we explored communicative language education and its central notion communicative competence. We started with a concise methodological history to see what we can learn from the past.

We contrasted the grammar-translation method with approaches that attempted to focus more on meaning or language use to varying degrees. The methods we subsequently dealt with were the direct method, reform method, audiolingual method and audiovisual method.

We characterised a direct method by its extensive and preferably exclusive use of L2. This conglomeration of approaches seems to hold the promise that learners learn to communicate more effectively when they are challenged to understand oral and written language and use the target language in meaningful ways. The question raised, however, was in how far the learners would be able to build up the grammar and vocabulary needed to communicate correctly and efficiently in another language, particularly if immersion in the language is only limited to two or three classes a week.

Next, we discussed four representatives of the so-called reform method, Viëtor, Sweet, Jespersen and Palmer. They united their specific interests in linguistics with their preferences for more meaningful L2 use. Viëtor stressed the importance of focusing on everyday living language. Sweet felt a focus on phonetics and pronunciation practice would help learners to communicate more correctly and effectively. Jespersen advocated the use of contextual materials and drills connected with meaningful L2 use. Finally, Palmer was in favour of an eclectic approach by merging insights from linguistics with understandings from pedagogy and psychology.

The audiolingual method was strongly influenced by behaviourism, structural linguistics, and the opportunities offered by audio-technology. It was discussed as a rather mechanical attempt to confront learners with the target language by way of stimulus-response structural drills. Audiolingualism is often contrasted with more communicative approaches.

Representatives of the audiovisual method all share an interest in a rich learning environment full of authentic oral and visual stimuli. We discussed a relatively influential and characteristically European example of audiovisualism, i.e. the structuroglobal audiovidual method (SGAV). The SGAV attempted to unite a focus on phonetics, grammar and vocabulary with a focus on meaningful communication. As such, it can be viewed as a prelude to communicative approaches.

We ended our methodological discussion by contrasting six basic tenets of audiolingualism with their counterparts in more communicative approaches.

Our methodological history has taught us about the importance of neither rejecting the past, nor of uncritically accepting the future as far as methodologies for the teaching and learning of another language are concerned. There is no such thing as the one and only successful method or approach to learn to communicate in a second or foreign language. What one can at best hope for is to retain what appears to have worked and develop what needs to be improved. In this study we do believe, though, in the potential of approaches that aim at meaningful communication, in which the knowledge components and skills that make up communicative competence are specified. Such knowledge and skills are needed in communication, which is, with Savignon (1997: 14), seen as “a continuous process of the *expression*, *interpretation*, and *negotiation* of meaning”. In the remainder of the chapter we defined communicative competence in more detail.

As an introduction to a specification of communicative competence, we referred to two theoretical issues, i.e. the view that language is increasingly seen as social behaviour and that language use can be characterised by a variety of functions. Both theoretical issues led to specifications of communicative competence. We then discussed two specifying models that have been influential over the years, that is Canale & Swain’s model (1980), later refined by Canale (1983), and Bachman’s model (1990), which was largely based on Canale & Swain’s. The knowledge and skills Canale and Swain originally identified have been represented in four competence areas: grammatical, discourse, sociolinguistic and strategic competence. The four areas represent neither hierarchy, nor sequence. All of these competences, together with the learner’s general knowledge, psychophysiological mechanisms and situational contexts (Bachman, 1990), make up communicative language ability. Whichever model one prefers, communicative competence is

generally seen to consist of a linguistic component, a pragmatics component and a strategic component.

We then dealt with three interpretations of communicative competence, i.e. *from surface structure to meaning*, *from meaning to surface structure* and *specification of context*. Advocates of a 'from surface structure to meaning approach' see skill-getting as the first and necessary step in language learning. They feel that learners will only be able to use the language correctly and appropriately if they have first acquired the necessary grammatical and idiomatic knowledge and skills. Those who favour a 'from meaning to surface structure' approach generally believe that grammatical and formal exercises should be generated by the needs that stem from communicative experiences. Language practice and discrete exercises are felt to be most useful when they accompany or follow rather than precede communicative experiences. An atomistic or sequential view of language learning is rejected. A final interpretation of communicative language ability is a 'specification of context' approach. Proponents of this approach are interested in selecting the language to which the learner is exposed in particular social contexts. Seminal work in defining social contexts of foreign language communication and subsequently developing assessment scales and descriptors has been done by the Council of Europe. The work culminated in the publication of *A Common European Framework* (CEF) *of Reference for Languages: learning, teaching and assessment* (2001). We claimed that the CEF can be seen as a document that further specifies the contexts offered by e.g. Canale & Swain's and Bachman's models of communicative competence in a detailed way.

To highlight consensus rather than controversy in foreign language education, we referred to eight aspects of the construct (Berns, 1990) on which there is general agreement.

We closed off the chapter with a review of research interests related to communicative language education characterised respectively by a focus on *meaning*, *form* and *cultural aspects of language use*.

As an example of research on meaning, we reported the successes of Canadian-French immersion programmes. They provided convincing empirical evidence for the positive effects of engaging classroom learners in meaningful L2 communication on their communicative competence. Despite the relative success of the immersion programmes, it was shown that comprehensible input was a necessary, but in itself an insufficient condition for language acquisition. It led to studies focusing on deficiencies in the input offered to the learners as well as to research into the absence of a focus on form in communicative classrooms. Around the same time, R.Ellis (1985, 1997) found that explicit grammatical instruction and error correction appeared to have very limited effects on the learners' communicative abilities. He discovered that grammatical structures appear to be acquired according to an immutable developmental sequence that is unaffected by instruction. Even though form-focused instruction is seen to be effective under certain conditions (R.Ellis, 2002), teachers cannot ignore the limitations of explicit ex-cathedra grammar teaching and error correction. Next, we put forward that studies were not limited to learner input, but focused on the importance of comprehensible output as well. Other studies stressed the importance of negotiated interaction and a host of studies focused on the significance as well as the nature of communicative tasks in successfully acquiring communicative language abilities.

As examples of research on form, we first referred to studies investigating what types of intervention may be effective and when. Important as they are, these studies mainly addressed the linguistic component of communicative competence. We then discussed Larsen-Freeman's (2003) approach to grammar in communicative language education, in which she attempts to bridge the gaps between linguistic competence and other competence areas that help learners learn how to communicate in another language. She coined the term *grammaring*, which refers to grammatical instruction as a 'process of doing grammar'. This process is meant to

result in learners using grammar structures accurately, meaningfully, and appropriately. According to Larsen-Freeman, learners learn to figure out their choices by passing through the three dimensions of form, meaning/semantics, and pragmatics/use.

Finally, we mentioned studies that highlight cultural aspects of language use. They were based on the belief that culture and language are intertwined. The assumption is that culture shapes and frames the human mind and affects what language is used and how it is used to interpret reality. We also referred to sociocultural studies on the nature of negotiated and socially-mediated interaction when learners learn another language. Concepts are embraced such as ‘collaborative dialogue’ and ‘languaging’ (Swain, 2006) and ‘action-based teaching, autonomy and identity’ (Van Lier, 2007).

The methodological history and the interpretations of communicative competence we discussed in this chapter will be used to analytically generalize how the three respondent teachers interpret communicative language education and the knowledge and skills they consider necessary for their learners to be able to communicate in English. We will report on this generalization in chapter 11 *Cross-Case Analyses*.

CHAPTER 5: FOREIGN LANGUAGE ASSESSMENT AND EVALUATION

5.1 Introduction

In the previous two chapters we explored *learner autonomy* and *communicative language education*. In this chapter we will examine the field of foreign language assessment and evaluation. We will explore the third and final construct central to our investigations in four sections.

First of all, we will indicate how testing, assessment and evaluation have been defined in the present study. There has not always been consensus on how the three notions are viewed in language testing literature and educational practice. Agreement on how these notions are defined is considered essential in the light of this study. In our investigation, we particularly focus on exploring the relationship between a language test and the teaching and learning process in settings that aim to develop learner autonomy as well as communicative competence.

Secondly, we will survey language testing from a historical perspective and report on four trends in language testing. The first three were mentioned by Spolsky (1978) as the *pre-scientific*, *psychometric-structuralist* and *integrative-sociolinguistic* trends. In addition, we will go into a fourth and more recent trend in language testing, which we have labelled as *critical-dynamic.* The four trends are linked with theories that were prevalent at the time a particular direction was favoured. Thus, the pre-scientific trend is characterised by a lack of theory. Psychometrics, American structuralism and behavourism influenced the psychometric-structuralist trend. The interest in language as social behaviour, language functions and communicative competence resulted in the growth of an integrative-sociolinguistic direction in language testing. Finally, philosophical postmodernism, critical theory, critical pedagogy, and dynamic language testing gave rise to an interest in testing ethics and more dynamic and interactive conceptions of learning, teaching and testing. It resulted in a more recent trend referred to as critical-dynamic. In discussing the four trends, a number of terms and notions of the field of language testing will be dealt with.

Thirdly, we will zoom in on a selection of professional standards in assessing and evaluating another language. In the first paragraph we will go into the importance of specifying *test purpose* and its related option for a particular *test type*. In the second paragraph, we will discuss the essential measurement qualities of *reliability and validity*, with particular attention paid to a unified conception of construct validity. Finally, we will discuss Bachman & Palmer's (1996) notion of *test usefulness*, which was based upon the model of communicative language ability they developed.

The final section of this chapter bears the title of the present study: *Testing for autonomy*. After a summary of the concerns discussed in this chapter, we will relate our theoretical explorations to what we discussed in the chapters on learner autonomy and communicative competence. After all, learner autonomy assumes content. In the case of foreign language teaching and learning, this content is increasingly determined by communicative curricula. The essential issue in our quest for autonomy is how the pedagogical goal of learner autonomy and the curricular goal of communicative competence are reflected in the respondents' assessment and evaluation practices.

5.2 Testing, assessment and evaluation

It is not always clear what is meant by testing, assessment and evaluation in the language testing literature and in testing practice. This is particularly true for distinctions between assessment and evaluation. This problem was tackled by members of the Association of Language Testers in Europe (ALTE). To achieve

consensus, a multilingual glossary of language testing terms was published in ten European languages (ALTE, 1998). The ALTE members felt it was vital to have equivalent terminology related to assessment and language testing in view of language testing training and research. In this paragraph, we will briefly discuss the definitions ALTE suggest for a test, assessment and evaluation.

A **test** is defined as follows:

> *A procedure for testing specific aspects of proficiency or knowledge.*
> *i. a set of components which together constitute an assessment procedure, often used to mean the same as an examination.*
> *ii. a single task or component for assessing an area of skill or knowledge, e.g. speaking or writing. In this sense a test may also form part of a complete examination as a component (e.g. the speaking test) or as a single task (e.g. cloze test).*
> iii. *an assessment procedure which is relatively short and easy to administer, often devised and administered within an institution (e.g. a progress test) or used as part of a research programme or for validation purposes (e.g. anchor test).*
> (1998: 166)

This broad definition covers any use of procedures in which specific aspects of proficiency or knowledge are tested. It may refer to a set of components, a single task or component, or to a practical assessment procedure. It may refer to an examination or part of an examination, giving the test taker some legal and/or social right if the results of a test lead to some formal qualification or certification. The definition offered for assessment is in line with the definition of a test.

> ***Assessment****. In language testing, the measurement of one or more aspects of language proficiency, by means of some form of test or procedure.*
> (1998: 135)

An assessment procedure is about the measurement of one or more aspects of language proficiency. ‘In language testing’ suggests that assessment is interpreted differently in contexts other than language testing. This is indeed the case. Two examples come to mind of the use of the notion of assessment in contexts different from language testing. Consider, e.g. career agencies or the military. Career agencies offer and perform assessments of a client’s interests, ability and capacities. Their ‘measuring’ procedure will almost always include aspects of evaluation. Inferences of test data are often made as part of the assessment procedure. Yet, we would like to argue that the interpretation of test results is an instance of evaluation rather than assessment. In assessment procedures like this, the distinction between assessment and evaluation tends to become blurred. Such conceptualisations of assessment outside language testing may, in their turn, influence the way how assessment and evaluation are interpreted in language testing literature and practice. The result is that assessment and evaluation are sometimes used interchangeably. It is not always clear whether a person refers to actual measurement procedures or to the inferences made from their results. A second example comes from the military. In this example, assessment and evaluation are less likely to get mixed up. When a reconnaissance soldier is asked to give an assessment of a war-like situation, (s)he is expected to report on the situation on the basis of observable data. Inferences or evaluations made on the basis of these observations are generally left to a superior. With tongue in cheek, we might observe that a foreign language learner is in a position not unlike our reconnaissance soldier. A language learner may also tend to safely leave the interpretation of test results to the teacher, who is assumed to have both the power and the expertise to make valid evaluations. The comparison begs

another question. One could wonder what observable data of language use a language teacher actually tests. After all, it generally is the task of a teacher of another language to assess the proficiency and communicative abilities of his/her learners and subsequently interpret or evaluate the results of the assessment. It is a concern we will follow up with the three respondent teachers in this study. How, then, do we distinguish between assessment and evaluation in this study? We will again look at the definition suggested by ALTE.

Evaluation differs from assessment in the eyes of the Association of Language Testers in Europe.

> ***Evaluation****. Gathering information with the intention of using it as a basis for decision-making. In language testing, evaluation may focus on the effectiveness or impact of a programme of instruction, examination or project.* (1998: 144)

According to ALTE, evaluation differs from assessment in its intention or objectives. Whereas assessment is primarily aimed at measurement at a given point in time, evaluation is about making inferences on the basis of information that has been gathered. Evaluation is a notion which may be directly related to some assessment procedure, but essentially differs from its goal.

In this study we follow the distinctions made by ALTE. This is how the notions of assessment and evaluation have been used in this investigation.

Assessment is considered as the classification of a learner's knowledge and skills at a certain point in time by some test or procedure.

Evaluation is seen as a retrospective and prospective procedure in which the results of an assessment are interpreted.

Assessment in itself does not necessarily lead to a learner's improved competence or performance. On the other hand, evaluation potentially has that effect. The assumption is that learners as well as teachers are more likely to be successful in the future if they are able and willing to critically evaluate the learning and teaching experiences that have led to a particular mark, score or grade. Future success in this study is seen in terms of a learner's degree of autonomy, level of language proficiency and effective assessment and evaluation practice

Having specified how the notions testing, assessment and evaluation are used in this study, we will now survey four trends in second and foreign language testing.

5.3 Four trends in second and foreign language testing

Spolsky (1978) distinguished three historical periods of modern language testing up to the time of his writing. We will use Spolsky's labels and subsequently discuss the *pre-scientific*, *psychometric-structuralist*, *integrative-sociolinguistic* trends. In addition we will go into some of the more recent developments and issues in the field of language testing. We will refer to these developments and issues as the *critical-dynamic* trend. The discussions will include explanations of well-known notions in the field of language testing, such as informal assessment, formal assessment, discrete-point and integrative tests, direct and indirect tests, and formative and summative assessment.

It is important to state that the four trends do not represent clearly defined historical periods in the sense that one period simply ends and is succeeded by another. Aspects commonly associated with one of the four trends may prevail in various combinations in everyday testing practice. We made similar remarks when

we discussed approaches or 'methods' in second or foreign language learning and teaching and in our discussion of important tenets of audiolingualism and more communicative approaches (see 4.1).

5.3.1 The pre-scientific trend

The pre-scientific trend generally refers to the period prior to the application of principles of educational psychology to language testing. (Farhady, 1980: 29). The assessment instruments developed in this period show no explicit concern with test properties such as reliability, objectivity or validity.

Language assessors following the pre-scientific trend typically favour a grammar-translation perspective on foreign language education. The major goal of the grammar-translation method is to teach the idiom and grammar of a language by way of translations. The oral skills of speaking and listening are generally ignored. Neither the approach to foreign language education, nor the tests themselves deal with language as communication. Preferred test types are translations of literary texts both to and from L1 and L2, dictation tests administered by word-by-word reading, and composition tasks in which the testees are tested for accuracy. Judgment and expertise are generally in a single assessor, who assesses mastery of vocabulary and grammatical knowledge globally, i.e. the assessor gives a single mark according to the impression made by the test taker's (re)production of knowledge and/or skills. Apart from idiom and grammar, the assessor hardly distinguishes between aspects of language use. "In its simplest form, it assumes that one can and must rely completely on the judgment of an experienced teacher, who can tell after a few minutes' conversation, or after reading a student's essay, what mark to give." (Spolsky, 1978:V).

Language testing in pre-scientific times may leave a predominantly negative impression. This impression should be toned down somewhat for three reasons. First, there have always been successful teachers, testers and tests, which includes pre-scientific times. In all likelihood, teachers such as Socrates and Comenius were quite capable of objective, reliable and valid assessments of their learners' abilities and achievements. In addition, there may have been numerous others, largely left anonymous by history. It would be wrong to conclude that objective, reliable and valid assessments and evaluations can only be made if the assessor is conscious of important test properties and notions. It could easily lead to the assumption that the judgments of teachers who are uninformed by testing theory tend to be subjective and unreliable and that the inferences they make on the basis of the test results are generally invalid. A second reason why the pre-scientific trend should not be exclusively seen in a negative light is the fact that informal assessment has always been part of social life to varying degrees. By *informal assessment* we mean judgments of situations or people based on some observable behaviour or product, without any conscious efforts to standardise the procedure and validate its findings. As such, informal assessment is not necessarily restricted to education. In language education, informal assessment usually concerns the testing of particular knowledge or skills. Informal tests and procedures are generally teacher-made. Thirdly, history has informed us about sophisticated forms of pre-scientific assessment. There are examples of large-scale language tests in the so-called pre-scientific "period". The first large-scale language tests were types of formal assessment. *Formal assessment* typically involves cooperation in test construction, administration and the interpretation of its results. Efforts are made to ensure that assessors adhere to an agreed procedure and apply rating procedures in appropriate ways. In *Measured Words* (Spolsky, 1995) the author provides more detail on early large-scale language tests. Kunnan (1999) summarized the instances as follows:

> *In the eighteenth and nineteenth centuries, formal examinations in Prussia, France, and Germany played a role in controlling schools and selecting civil*

> *servants, and in Britain, Oxford and Cambridge Universities started examinations with public oral disputations and then added written examinations. In the US, written examinations were introduced at Harvard University in the mid-nineteenth century. In the first half of the twentieth century, language tests gained in stature in the US as the Army Alpha tests wee instituted after the US entered World War I and Cambridge University in Britain began to administer the Certificate of Proficiency and the First Certificate tests in English.*
> (Kunnan, 1999: 707)

Earlier examples of formal assessment of proficiency can be found as well. In Europe, the medieval guilds tested the performance of their trainees by means of so-called *Proeven van bekwaamheid* (Practical tests of ability and performance). Elsewhere, similar types of performance assessment may have existed.

Before discussing the psychometric-structuralist trend, we will turn to an example of one of the oldest language performance tests known to mankind. It is the example of the so-called *Shibboleth* test, taken from, perhaps quite tellingly in view of our interest in assessment and evaluation, the Book of Judges in the Old Testament. Failure on the test had rather drastic consequences for the test taker. The test was first elaborated on by Spolsky (1995) and later used in an address by McNamara (2002). The illustration shows how past developments are linked with more recent issues in foreign language testing. McNamara used the example of the Shibboleth test in an invited address for the Language Assessment Ethics Conference in Pasadena, CA on May 17, 2002. As an illustration, part of the introduction of McNamara's address has been quoted below:

> *The title of this talk comes from an incident in the Old Testament during a time of conflict. It refers to a language test used to distinguish members of two warring tribes, who spoke essentially the same language but had small differences in pronunciation. A bit like Kiwis and Aussies: imagine the Ansett/Air New Zealand conflict taken to lethal proportions and Kiwis in Australia trying to blend in with the Australian population to avoid vengeance of Australians. Here is the passage from the Bible: for Gilead, read Australia; for Ephraim, read New Zealand:*
>
> *Judges, 12 4-6) (The Jerusalem Bible translation)*
>
> *4 Then Jephthah mustered all the men of Gilead and joined battle with Ephraim, and the men of Gilead routed Ephraim…*
>
> *5 Then Gilead cut Ephraim off from the fords of the Jordan, and whenever an Ephraimite fugitive said: "Let me cross," the men of Gilead asked him, "Are you an Ephraimite?" If he answered "No"*
>
> *6 they said, "Then say Shibboleth." He would say "Sibboleth," since he could not pronounce the word correctly. Thereupon they seized and slaughtered him by the fords of the Jordan. There perished in this way forty-two thousand men of Ephraim.*
>
> *In a note to verse 6, The New American Bible says:*
>
> *Shibboleth: "an ear of grain." But this Hebrew word can also mean "flood water" as in psalm 69, 3, 16. Apparently, the Gileadites engaged the Ephraimites in conversation about the "flood water" of the Jordan. Differences in enunciating the initial sibilant of the Hebrew word betrayed different tribal affinities.*

> *So this test was what we would call a performance test, where the test is naturalized and disguised in the context of an apparently casual conversation.*
> (McNamara, 2002)

The quotation above is more than an illustration of a large-scale practical language test from pre-scientific times, with its rather drastic impact on 42,000 unfortunate test takers. McNamara's reference to the Shibboleth test in an address on a conference on language ethics is an illustration of how the past relates to the present in recent approaches of language testing. The interest in language ethics is an instance of a more recent issue in second and foreign language testing, which we will discuss as one of the developments in the trend we labelled as *critical-dynamic*. Yet, we will first elaborate on a period when the field of language testing was invaded by experts.

5.3.2 The psychometric-structuralist trend

Testing theory is relatively young. Its start is often associated with the works of psychologist Cronbach and linguist Lado in the late 1950s and early 1960s. Cronbach and Lado were among the leading theorists behind types of second and foreign language assessment in the psychometric-structuralist trend. Spolsky (1995) reports that the trends were set in the 1940s, when the first two Ph.D. studies on language testing were conducted. He mentions Villareal's *Test of Aural Comprehension* (1947) and Lado's *Measurement in English as a Foreign Language* (1949). In addition, psychologist Cronbach published his widely-cited measurement paper 'Coefficient alpha and the internal structure of tests' (1951). Cronbach was a scientist and educational psychologist interested in psychometrics, that is in the measurement of mental abilities and processes, such as language proficiency. As an educational psychologist, Cronbach was familiar with behaviourism in all its forms, from Watson's classic behaviourism to Skinner's more enlightened operant variety.

Developments in psychometrics and psychology continued to be important to the field of foreign language testing. In 1954 the first *Standards for Educational and Psychological Tests and Manuals* was published by the American Psychological Association (APA). The standard outlined five types of test validity, a test quality determined by the question whether an assessment procedure actually assesses what it intends to assess. The five types of validity were face-content validity, criterion-related validity, predictive validity, concurrent validity and construct validity. These types of validity will be discussed in more detail later in this chapter.

Vilareal and Lado were linguists interested in language testing. They were influenced by American Structuralist views of language. Structural linguists saw language as consisting of separate items that could be arranged in a variety of ways (Spolsky, 1978:6). These ideas matched well with the behaviourist learning theories that prevailed in the first half of the twentieth century and beyond. Psychometrics and structuralism found each other in a happy marriage. The psychological theories and practical needs of testers fitted in well with American structuralist views of language. The structuralists knew how to sample long lists of small items by way of contrastive analysis. These separate items, such as pronunciation, vocabulary and a wide variety of grammatical items, became known as *discrete points.* The name of Lado is usually associated with discrete-point testing, and he was soon considered to be the second and foreign language testing expert in the late 1950s and early 1960s. (Lado, 1957; 1961). These alleged components of language proficiency could be tested objectively. The psychometric experts knew how to provide "objective" measures, using various statistical techniques to assure reliability and certain kinds of validity, such as content, concurrent and predictive validity. They developed short-item, multiple-choice tests of the discrete items provided. These tests typically consisted of *indirect test items*. An *indirect test* is a test that attempts to measure the alleged

abilities underlying a language skill, rather than testing performance of the skill itself. It does not simulate any real world language activity or any aspect of one. An example would be a test of writing ability, in which a candidate is asked to mark structures that are used incorrectly in a text. In case of a *direct writing test* the performance of writing is directly tested, e.g. the test taker is asked to actually write a letter.

Behaviourism, psychometrics and American structuralism are primarily based on positivist views of human knowledge and learning. Positivists believe that true knowledge can be achieved, and can thus be identified and measured objectively, reliably and validly by way of quantitative procedures. The theory was first associated with the French philosopher Auguste Comte (1798 – 1857). In the 1920s and 1930s, logical positivism developed. Bertrand Russell and Ludwig Wittgenstein are representatives of this doctrine. They claimed that propositions are only meaningful if they can be verified empirically. Therefore, they considered aesthetics, religion and metaphysics as meaningless. This positivist paradigm influenced and dominated language testing and language testing research in the psychometric-structuralist trend.

The interest in positivism, psychometrics and structuralism in second and foreign testing, with its emphasis on observable and measurable data, provided academic status to the field. Attempts were made to understand and identify components of language proficiency. Lado's efforts (1961) have proved to be seminal in that respect. He developed a 'skills and components' model, which was to influence language teaching, learning and testing for years to come. A lot of teachers, testers and researchers were inspired or influenced by Lado's ideas. Kunnan (1999) summarizes the model as follows:

> *…the 'skills and components' model, was proposed by Lado (1961) in a grid form with language skills (listening, speaking, reading, and writing) crossed with language knowledge and components (phonology/graphology, structure, vocabulary, rate, and general fluency). The 4 x 4 grid's 16 different cells (see Harris, 1969) were considered independent; thus requiring test developers to construct 16 separate tests for full coverage. The Michigan Test of English Language Proficiency and the early versions of the Test of English as a Foreign Language are the best examples of tests that followed this model.*
> (Kunnan, 1999: 707/8)

Just as 'pre-scientific' forms of second or foreign language learning, teaching and testing are likely to linger on, there may still be evidence of the psychometric-structuralist trend in a school teacher's assessment and evaluation practice. We will identify and discuss these in more detail in chapter 11, when we analyse the data of the language testing practice of our three respondent teachers. The final period Spolsky (1978) identified was the integrative-sociolinguistic period or trend.

5.3.3 The integrative-sociolinguistic trend

The integrative-sociolinguistic trend derives its name from integrated forms of language testing, known as integrative tests, and from theoretical advances in the field of sociolinguistics, most notably the interest in language as social behaviour and in language functions we discussed in chapter 4.

Integrative testing

Unlike discrete-point testing, integrative testing involves items or tasks that require more than one skill or subskill to complete. A cloze test is an example of an integrative test. It is a type of gap-filling task in which whole words are deleted from a text. The original words, or acceptable synonyms, have to be supplied by the test taker. Other Examples of integrative tests are oral interviews, or tasks such as

reading an article and writing a written response to it. The name of Carroll is generally associated with these types of test.

As early as 1961, Carroll attempted to expand Lado's skills and components model. He was among the first scholars to call for *integrative testing.* It refers to assessment procedures or test items in which knowledge of various components of a language system has to be integrated with an ability to produce and interpret language appropriately in its linguistic context (McNamara, 2000:133). Integrative testing taps into a number of abilities, such as the grammatical, textual or rhetorical competences required to supply a cloze test item. (Chapelle & Abraham, 1990: 121). Carroll (1961) acknowledged that Lado's approach was well-grounded in theory, but suggested to supplement the model with tests measuring speaking, writing and listening skills.

Carroll's suggestion reflects the audio-lingual and communicative perspectives on second and foreign language education discussed in the previous chapter**.** Gradually, more and more theorists, researchers and practitioners became convinced that language had to be studied in its social context.

The second part of the integrative-sociolinguistic label derives its name from a branch of anthropological linguistics called sociolinguistics, which studies relationships between language and society. Sociolinguists investigate how language and culture are related and how language is used in different social contexts. The field has many 'founding fathers', such as Durkheim, Labov, Mead, Meillet, Sapir etc. Hymes is generally considered a founder of sociolinguistics as it emerged in the 1960s and 1970s. We already discussed that his theoretical formulation of communicative competence, which Hymes formulated as distinct from Chomsky's linguistic competence, was to have a lasting influence on second and foreign language teaching. Hymes also influenced language testing.

Hymes' formulation of communicative competence and performance (1972) was followed by a host of labels referring to tests of communicative ability: 'direct', 'functional', 'authentic', and 'performance'. Kunnan (1999: 707) claims that it was the notion of performance testing that particularly emerged from all of the other labels. In this chapter we already referred to the Shibboleth test as one of the first performance tests in history. A performance test is a test procedure which requires the candidate to produce a sample of language, either in writing or in speech (e.g. in essays or oral interviews). Such procedures are often designed to replicate language performance in non-test contexts. The assessment procedures used may be as varied as open-ended constructed responses, problem-solving and essay tasks, portfolios, and simulations of real world problems. (Shohamy, 1995).

Sociolinguistic aspects of language testing

Sociolinguistics inspired teachers and researchers to adopt a more communicative perspective on second and foreign language learning, teaching and testing. Present research areas of sociolinguistics can be as diverse as language and social interaction, language contact and change, language and social issues (e.g. immigration, unemployment, discrimination), mediated communication, language ideology, discourse and grammar, intercultural and crosscultural communication, gender and language, narrative and identity, language and aging, endangered languages, language and social theory, and language and the professions (e.g. law, medicine, journalism, education).

Gradually, theorists worked towards supplements and alternatives to Lado's skills and components model. Kunnan (1999) singles out two major models of language proficiency: Oller's pragmatic model (1979) and the communicative models of Canale and Swain (1980) and Bachman (1990), later elaborated by Bachman & Palmer (1996). These models were embraced by advocates of the integrative-sociolinguistic trend in second and foreign language testing, and gave way to new types of test and assessment procedures.

Towards communicative models

According to Kunnan (1999), Oller was the first to actually promote integrative testing and develop an alternative to Lado's skills and components model. To counter the problems with discrete-point tests, he proposed a *pragmatic test model.* Oller defined language tests not in terms of the discrete elements of knowledge to be tested, 'but in terms of the language processing operations required of learners'. (McNamara, 2000: 92). Oller defined pragmatic tests as:

> *'…any procedure or task that causes the learner to process sequences of elements in a language that conform to the normal contextual constraints of that language and which require the learner to relate sequences of linguistic elements via pragmatic mapping to extralinguistic context.*
> (Oller, 1979: 38)

Discrete-point items, such as the manipulation of sequences of verbal elements, should be tested by using them in a variety of contexts in which they are normally used. A test, and the contexts provided by the test, should invoke and challenge a learner's developing grammatical system. In addition, a test should cause the test taker to link linguistic knowledge to extralinguistic contexts in meaningful ways. Extralinguistic contexts invite the language user to pay attention to sociolinguistic conventions, such as markers of social relations, politeness conventions, expressions of folk-wisdom, register differences and dialect and accent (CEF, 1998). Pragmatic competence is a common component in models of communicative language ability. It is concerned with the user or learner's knowledge of the underlying principles according to which messages are organized, structured and arranged ('discourse competence'), used to perform communicative functions ('functional competence') and sequenced according to interactional and transactional schemata, i.e. to patterns of social interaction. Oller strongly objected to test tasks in which separate items are tested without any meaningful context. He objected to all discrete point tests, the rote recital of sequences of material without attention to meaning, the manipulation of verbal elements, possibly in complex ways, but in ways that do not require awareness of meaning.

Oller first promoted dictation and cloze procedures as examples of pragmatic tests. He argued that cloze tests in particular were economical and efficient tests that demanded the test taker to integrate grammatical, lexical, contextual and pragmatic knowledge. Cloze tests are reading tests that generally consist of about 400 words, in which, after a few introductory sentences, every 5^{th} to 15^{th} word has been removed and replaced with blanks (Oller, 1979:365). The task for the test taker is to supply this missing words, either the exact word replacement or any acceptable word replacement. Cloze tests were easy to construct and relatively easy to score and were at first seen as attractive alternatives to more elaborate and expensive tests of the productive skills of speaking and writing (McNamara, 2000:16). Later, when it was shown that cloze tests seemed to be measuring the same kinds of knowledge and abilities as discrete point tests of grammar and vocabulary, Oller focused more on productive oral communication, essays, and writing tasks that met his criteria of pragmatic tests.

Along with his conception of pragmatic tests, Oller hypothesized that language proficiency is unitary and not divisible, as discrete-point testers had suggested in the past. He posited the existence of a single competence underlying all language skills. After much debate, a weaker version of Oller's unitary competence hypothesis was supported by most researchers. According to Kunnan (1999):

> *This middle-ground position claims a multicomponential view of language proficiency, that there is a general component or variance common to all*

> *aspects of language proficiency and that a smaller component or variance is unique and shared only by some aspects of language proficiency.*
> (1999: 708).

The next models of language proficiency Kunnan discusses are the communicative models introduced by Canale and Swain (1980), and later developed by Bachman (1990) and Bachman and Palmer (1996). In the previous chapter we already referred to the Canale & Swain's and Bachman's models, with its division of communicative competence into grammatical, discourse, sociolinguistic and strategic competences. These models do not only specify communicative language ability, but may also facilitate tests of communicative competence. Below, partly as a reminder, we will present Bachman & Palmer's (1996) model. They distinguished the following areas of language knowledge:

> ***Organizational knowledge***
> *(How sentences and utterances are organized)*
> ***Grammatical knowledge***
> *(How individual utterances or sentences are organized)*
> *Knowledge of vocabulary*
> *Knowledge of syntax*
> *Knowledge of phonology/graphology*
> ***Textual knowledge***
> *(How utterances or sentences are organized to form texts)*
> *Knowledge of cohesion*
> *Knowledge of rhetorical or conversational organization*
> ***Pragmatic knowledge***
> *(How utterances or sentences and texts are related to the communicative goals of the language user and to the features of the language use setting.)*
> ***Functional knowledge***
> *(How utterances or sentences and texts are related to the communicative goals of language users)*
> *Knowledge of ideational functions*
> *Knowledge of manipulative functions*
> *Knowledge of heuristic functions*
> *Knowledge of imaginative functions*
> ***Sociolinguistic knowledge***
> *(How utterances or sentences and texts are related to features of the language use setting.)*
> *Knowledge of dialects/varieties*
> *Knowledge of registers*
> *Knowledge of natural or idiomatic expressions*
> *Knowledge of cultural references and figures of speech.*
> (Bachman and Palmer, 1996)

Bachman and Palmer's model is a conceptual model that does not have empirical support for all of the areas and components as yet. 'It is open for researchers to refute, refine or accept based on logical or empirical research' (Kunnan, 1999: 708).

Nevertheless, the new communicative theories of language and language use and their corresponding communicative models inspired the development of effective and useful tests that aimed to measure communicative ability. Tests gradually became more communicative.

Towards communicative tests

McNamara (2000) states that it took about a decade before Hymes's theory on language and language use actually impacted language testing practice. He claims that communicative language tests ultimately came to have two features:

> 1. *They were performance tests, requiring assessment to be carried out when the learner or candidate was engaged in an extended act of communication, either receptive or productive, or both.*
> 2. *They paid attention to the social roles candidates were likely to assume in real world settings, and offered a means of specifying the demands of such roles in detail.*
>
> (2000:16/7)

It is feature number two in particular that distinguishes communicative language tests from the original integrative/pragmatic tests inspired by Oller's Unitary Competence Hypothesis. The focus on authentic texts and real world tasks came to be essential characteristics of tests for communicative ability. McNamara (2000:17) mentions that developments in Britain were most notable in that respect, e.g. in the tests developed by the Royal Society of Arts, the British council and other authorities that developed communicative tests of English as a foreign language.

The notion of authenticity in second and foreign language teaching and testing has been explored and debated (Breen, 1985; Widdowson, 1990; Kramsch, 1993; Taylor 1994). The debates have often been caused by a failure to recognize the several types of authenticity that come to the fore when another language is learned. Quoting Breen (1965), Taylor (1994: 1-2) distinguishes four types of authenticity:

> *1 Authenticity of the texts which we may use as input data for our learners.*
> *2 Authenticity of the learners' own interpretation of such texts.*
> *3 Authenticity of the tasks conducive to language learning.*
> *4 Authenticity of the actual social situation of the language classroom.*
>
> (Breen, 1985: 61)

According to Taylor, it is too readily assumed that there is some sort of global and absolute notion of authenticity in which all the different kinds must be simultaneously and completely present. While there are relatively clear definitions available of what is meant by authenticity in relation to teaching materials and texts, there is much less agreement about what constitutes authenticity of context and of task or activity. Moreover, a crucial point that is often overlooked is that the classroom has its own authenticity. For most learners the classroom is a very real and authentic place, despite the artificial and contrived nature of some of the exercises and tasks the learners are confronted with. Taylor (1994) does not see the so-called artificiality as a problem. He feels that:

> *…learners, in their capacity as knowers and users of language, are quite capable of extrapolating from the classroom situation, and that consequently we need not be worried about the so-called artificiality of the classroom situation. We need to remember that the language classroom is there to promote language learning. If we take activities where it is said that there is genuine communication, for example information gap type exercises, they are authentic in one sense, in that genuine communication take place, but the whole thing is still contrived, in the sense that it is aimed at language learning. This does not matter because the learners have the sense to know what is going on. They are used to the classroom situation and to the kind of activities that go on there. They can distinguish between skill-getting and skill-using, as we have already noted (and even recognise that these can go on simultaneously), they can create their own authenticity in the classroom*
> (1994:7-8).

We will return to the notion of authenticity in language testing, when we discuss a definition of test usefulness, which derives from Bachman & Palmer's (1996) model of communicative competence.

This ends our discussion of the integrative-sociolinguistic trend in language testing.

Some of the more recent concerns in language testing are not always satisfactorily covered by the labels we have used so far. In the next section, we will deal with two approaches to second and foreign language testing that offer perspectives that fundamentally differ from the ones we have explored so far. *Critical language testing* and *dynamic assessment* have lent their names to what we will discuss as the critical-dynamic trend in language testing. The two approaches are instances of more recent concerns in the field of language assessment and evaluation.

5.3.4 The critical-dynamic trend

Critical language testing and dynamic assessment are approaches to language testing characteristic of post-modern thought. By way of introduction to the critical-dynamic trend, we will briefly refer to philosophical postmodernism and the ways in which it has affected the field of language testing. After that, we will discuss critical language testing and dynamic assessment as approaches typical of postmodernism. The two approaches share a critical attitude towards existing language testing practices, which are viewed as rather static.

Post-modern language testing

Modernism attempted to describe the world in rational, empirical and objective terms. Modernist approaches assume that truths can objectively be uncovered in the everyday realities that make up the human condition. Intellectually, one could say that modernism started with the Enlightenment.

Philosophical postmodernists, such as Lyotard, Derrida, Foucault and Rorty, question the underlying certainties that are promised by reason. Postmodernism represents partly overlapping critical attitudes towards reality, change and difference, metaphysics, the self, inquiry and forms of scholarship. Philosophical postmodernism does not represent a single point of view, but is generally characterised by features such as ‘the challenging of convention, the mixing of styles, tolerance of ambiguity, emphasis on diversity, acceptance (indeed celebration) of innovation and change, and stress on the constructedness of reality.’ (Beck, 1993). Postmodernism has come to permeate architecture, the graphic arts, dance, music, literature, literary theory, and, indeed, the field of language testing.

According to Spolsky (1999), language testers have started to question the rigidity in approaches to second or foreign language assessment and evaluation in the approach to the new millennium. This rigidity concerned the practices of intuitive examiners as well as the works of testing professionals. On the one hand the informal practices of intuitive teachers were questioned. Largely uninformed by testing theory, teachers were often felt to produce tests of questionable reliability and superficial validity. On the other hand, language testers were critical of the authoritative and at times authoritarian psychometrist, whose reliable tests seemed to reduce the multidimensionality of communicative language ability to unidimensionality (1999: 702). Postmodern language testing aims at taking the best of both worlds.

> *It starts with a recognition of the ethical requirements of testing, the need for testers to share with test users responsibility for the kinds of decisions made with test results (Spolsky 1981, 1984). It accepts the need for careful analysis of the specific purposes of a test, and for the design of tests to meet these specifications. It accepts the need to make procedures of test preparation, scoring, and interpretation fully explicit and open to inspection by test takers and users. It recognizes the value of professional standards for testing. It encourages the combination of discrete-point and integrative*

> *tests. Most of all, it recognizes the limitations of using a single method for foreign language testing, or of relying for important decisions on a single test.*
> (Spolsky, 1999: 702)

Spolsky mentions five concerns that should be on the language tester's agenda towards the new millennium: (1) an interest in language testing ethics, (2) detailed attention to test purpose and design, (3) the need for test procedures to be open to inspection and feedback, (4) a recognition of professional standards, and finally (5) a preference for multiple methods of assessment and evaluation, so that important decisions do not depend on the results of a single test or type of test.

The challenges outlined above have partly been taken up by two approaches to language testing that offer a different perspective to assessment and evaluation than the ones we have discussed so far. We will subsequently discuss critical language testing and dynamic assessment.

Critical language testing

Today, tests are more and more seen as power tools, embedded in social and political contexts and agendas, related to intentions, effects and consequences and open to interpretations and values. This view has led to the need to critically examine the uses and consequences of tests and 'to monitor their power, minimize their detrimental force, reveal the misuses, and empower the test takers' (Shohamy, 2001:131). It has led to approaches that are generally referred to as critical language testing. The critical language testers' concerns are closely linked with postmodernism and the social themes raised by critical theory and pedagogy, constructivism, and sociocultural theory, which we discussed in chapter three on learner autonomy.

The critical language tester believes 'that the principles and practices that have become established as common sense or common knowledge are actually ideologically loaded to favour those in power, and so need to be exposed as an imposition on the powerless' (Mc Namara, 2000: 76). Language tests are seen as instruments of power that must be analysed critically. This is particularly true for 'industrialised" language tests, such as TOEFL, that are used on a large international scale. The claim to expertise is always a relative one to the critical language tester. Some critical language testers argue for dismissal of testing altogether. However, this is not a position taken by all.

Shohamy (2001:133/4) puts forward that critical language testing needs to ask critical questions that are meant to identify, often hidden, power dimensions and to expose unfairness. Examples of such questions are: Who are the testers? What is their agenda? Who are the test takers? What is their context? What is the context of the topic being tested? Who is going to benefit from the tests? Why is the test being given? What will its results be used for? What is being tested and why? What is not being tested and why? What are the underlying values behind the test? What are the testing methods? What additional evidence is collected about the topic? What kind of decisions are reached based on the test? Who, excluding the tester, is included in the design of the tests and its implementation? What ideology is delivered through the test? What are the messages about students, teachers and society that the test assumes? What type of feedback is being provided based on the test, and to whom? Can the test, its rationale and results be challenged? What are the intended and the unintended uses of the test? What are its washback effects? What are some of the ways that test takers and others use to challenge tests?

By seeking answers to the questions above, critical language testing attempts to democratise the testing process. To do so, critical testers advocate collaborative and dialogical approaches to assessment, highlight the social and ethical responsibility of the testers, and stress the rights of the test taker. Drawing on Shohamy (2001), we will briefly go into these areas of critical language testing in the next sections.

Collaborative and dialogical approaches to assessment

For a long time, assessment has been 'practised as a reflection of knowledge and values of those in authority, with very little attention given to the other interested agents, test takers included' (Shohamy, 2001:135). Assessors and evaluators, however, should be aware of their fallibility and compensate for limitations by working together with other agents that can provide evidence of assessment knowledge. In order to democratise assessment and evaluation practices, they should be made more collaborative and dialogical, with explicit roles for the test takers in particular. According to Darling-Hammond (1994), there is a democratic need to change the ways in which assessment and evaluation are used: from sorting mechanisms to diagnostic supports; from external monitors to locally generated tools for enquiring deeply into teaching and learning; and from purveyors of sanctions to levers for equalizing force.

Recent views of language testing, therefore, see the act of testing as a mutual effort of testers and test takers, along with other sources of knowledge, such as parents, teachers or peers. Through constructive, interpretive and dialogical sessions each participant collects language data and demonstrates it in a contextualised and responsible manner (Shohamy, 2001: 137). Such a view of assessment and evaluation involves the acceptance of interpretive research paradigms, which are different from the naturalist conception of social science, from which educational measurement has evolved.

Shohamy (1995) reports on a democratic assessment model that was meant to assess the language proficiency of immigrant students acquiring a new language. In that model, the students' language proficiency is assessed by teachers, the students themselves, and by the use of a standardised diagnostic test administered by a central body. The teachers collect data with their own tests and observations. The students provide evidence of their language performance through self-assessment and portfolios. All of the evidence gathered is subsequently discussed in an assessment conference of language teachers, classroom teachers, learners, and occasionally parents. Together, they discuss and interpret the information obtained and attempt to arrive at meaningful recommendations for pedagogical strategies of language improvement. Multiple sources of evidence and constructive feedback are hallmarks of models such as Shohamy's.

Social and ethical responsibilities of testers

It has increasingly become recognised that language testers, that is all those who have made some contribution to the act of testing, have professional and moral obligations to the test takers and to society as a whole (Davies, 1997; Spolsky, 1998, 1999). Davies (1997) observes that the morality of language testing is increasingly discussed both within and outside the field. He remarks that:

> *While the growing professionalization of language testing is perceived as a strength and a major contribution to a growing sense of ethicality, the increase in commercial and market forces, as well as the widespread use of language assessment as an instrument in government policy, may pressure language testers into dangerous (and unethical) conduct*
> (1997:236).

Spolsky (1998:13) comments on a similar phenomenon in his statement that the real-world desires for easy answers set major ethical challenges to testers who know the complexity of the task of assessing language proficiency. Davies (1997) argues in favour of an explicit statement of ethical conduct and practice for language testers in order to prevent conflicts that result from topics as varied as the relationship with stakeholders, the relationship between bias and fairness, washback, the politics of gate-keeping and a conflict between fairness and face validity. Shohamy (2001)

sums up some of the moral dilemmas language testers face as their professional roles and responsibilities, once they become aware of the misuses of tests in society.

> *Is the test developer responsible for uses and misuses of the test? What is the role of the tester once he or she notices misuses? Is the tester's role at that point to warn against misuses or actually take steps such as using sanctions against misuses? Or perhaps the tester has no responsibility to worry about the test takers after a test had been handed to the users*
> (2001: 145).

Views of the tester's responsibility vary. Shohamy (2001) identifies five perspectives. The first is Davies's (1997) argument for an ethical perspective of professional morality which is defined as a contract for the profession and the individual with the public, thereby safeguarding all three. However, Davies states that such a contract inevitably involves dilemmas, such as balancing one's professional conduct with one's moral conscience and may go as far as questioning the right of the profession to exist. Davies maintains that it is impossible for language testers to take account of all possible individual and social consequences. What they can at best do is to carry out an internal (technical) bias analysis and be willing to be accountable for a test's fairness (Davies, 1997:336).

Another view complements rather than contradicts Davies's argument. It is the view that the language tester is responsible for being open and explicit about the intentions, effects and consequences of tests. All of the information on the tests should be made accessible to the users and to society at large. Spolsky (1998) notes that the danger of tests comes in their misuses, which are less likely to occur if the inevitable uncertainty of a test is studied and publicised, with attention paid to the user's perspective.

A different view is expressed by Hamp-Lyons (1997). She argues that the language tester needs to accept responsibility for all consequences of which (s)he is aware.

> *The responsibility of the language testers is clear: we must accept responsibility for all those consequences which we are aware of. Furthermore, there needs to be a set of conditions and parameters inside which we are sure of the consequences of our work and we need to develop a conscious agenda to push outward the boundaries of our knowledge of the consequences of language tests and their practices*
> (1997:302)

The fourth view Shohamy discusses is more radical and particularly relates to professional high stakes language tests. According to this argument, constructing a test is identical to the manufacturing of other products in society. This makes the tester's responsibility comparable to the liability of a manufacturer in producing defective articles or engaging in malpractice. The responsibility of the tester, therefore, is not simply to be open about a test's intentions and effects, but also to impose sanctions, punish and forbid the use of tests that are misused and penalise those who violate the standard of correct testing practices.

The final view Shohamy goes into is based on the recognition of shared responsibility and shared discourse of the tester, user, and test taker when language ability is assessed and evaluated. The responsibilities for good conduct are in the hands of all those who are involved in the testing process. Shared responsibility starts with the testers' awareness of their limitations and their willingness to share their authority. In this view, testers realise that they cannot cover all domains of knowledge and skills, and that they do not have all the answers when undertaking assessment. Testers and test takers should therefore engage in a mutually constructive effort, working together in constructing the meaning of knowledge. Such a view relates to constructivism as well as to critical language testing. This

perspective merges some of the views on the tester's responsibility discussed above and involves:

> *the need to be critical about tests and their uses, to collect data on the effects and the consequences of tests, to warn against misuses, and to protect all those involved in the act of testing – testers as well as test takers. It should be viewed as the responsibility of all to work for better instruments and uses of tests and to warrant against misuses*
> (Shohamy 2001:148)

Related to the concern of shared responsibility is the issue of shared discourse and for specifying the type of information that needs to be collected. To effectively share information 'It is important to acquaint all those who have concerns about tests and their use with the techniques and vocabulary of the testing community so that they can enter into the discourse without being dismissed as naïve' (Shohamy, 2001:149). We will discuss test qualities, most notably the notion of validity, in more detail in the final section of this chapter.

The interest in professional morality in the field of language testing, has led to the development of standards. In 1981, the *Standards for Evaluation of Educational Programs, Projects and Materials* were developed. These standards had been developed over a number of years by the Joint Committee on Standards for Educational Evaluation, made up of members from the AERA, the APA, the NCME and nine other organisations. Nevo and Shohamy subsequently extended these standards to testing methods. The original 30 standards were reworded and reduced to 23. The standards were organised as follows:

1. ***Utility standards*** *The utility standards are intended to ensure that a testing method will serve the practical information needs of given audiences.*
   Utility assures that a test serves the practical information needs of a given audience; for example, it is concerned with the impact that tests may have on instruction and learning in the classroom, such as backwash effects. Also included are other utility factors such as finding the best ways for training testers and raters for presenting test results. they are also concerned with tester credibility, information scope, justified criteria, dissemnination timelines and impact.

2. ***Accuracy standards*** *The accuracy standards are intended 'to ensure that a testing method will reveal and convey technically adequate information on the educational achievement of those that are being tested'.*
   Thus accuracy assures that the test represents reliable and valid measurements, testing conditions, data analysis and objective reporting.

3. ***Feasibility standards*** *Feasibility standards are intended 'to ensure that a testing method will be realistic, prudent and frugal'. Feasibility relates to whether the tests are feasible to administer within different contexts, whether they are conducted with the support of certain political groups, whether they produce information of sufficient value to justify their costs, and whether it is practical to train people to conduct such tests.*

4. ***Fairness Standards*** *These are intended 'to ensure that a testing method is conducted legally, ethically, and with due regard to the welfare of tested individuals as well as those affected by test results'.*

> *Fairness relates to whether the tests are conducted legally and ethically, as well as with care for the welfare of the respondents. Included are questions on whether the tests are based on material that examinees are expected to know, if they are conducted so that the rights and welfare of human subjects are respected, if they are conducted fairly with regard to the strength and weaknesses of the individuals tested, and if the test takers have a positive attitude towards the tests.*
>
> (Shohamy, 2001:152)

Recently, the need to for a code of practice that will consider not only the traits of a test, but also its uses and consequences, has been acknowledged by the testing community. the International Language Testing Association (ILTA) developed a Code of Ethics, which was adopted in Vancouver in March 2000 (ILTA, 2000). The Code of Ethics identifies nine fundamental principles, each elaborated on by a series of annotations which are meant to clarify the principles and prescribe what ILTA members ought to do and not do. Failure to uphold the code of Ethics may have serious consequences for ILTA members, such as the withdrawal of membership. As an example of the Code, we will present the fifth principle, which directly relates to one of the objectives of the present study.

> ***Principle 5***
>
> *Language testers shall continue to develop their professional knowledge, sharing this knowledge with colleagues and other language professionals.*
>
> *Annotation*
>
> - *Continued learning and advancing one's knowledge are fundamental to the professional role; failure to do so constitutes a disservice to test takers.*
> - *Language testers shall make use of the various methods of continuing education that are available to them. These may involve participation in continuing language testing programmes.*
> - *Language testers shall take the opportunity to interact with colleagues and other relevant language professionals as an important means of developing their professional knowledge.*
> - *Language testers shall share new knowledge with colleagues by publication in recognized professional journals or at meetings.*
> - *Language testers shall be expected to contribute to the education and professional development of language testers in training and to the drawing up of guidelines for the core requirements of that training.*
> - *Language testers shall be prepared to contribute to the education of students in the wider language professions.*
>
> (ILTA, 2000)

The principle above illustrates the commitment of ILTA members to the issue of professional morality, in particular their dedication to professional development and the dissemination of knowledge, views and experiences. It highlights the need for collaboration and dialogue of professional language testers with other professionals involved in language education, such as teachers and teachers-to-be. One of the initiatives worth mentioning that attempt to bridge any gaps between testing and teaching professionals and set up constructive dialogue was taken by CITO, an institute for educational measurement in the Netherlands. CITO was founded in 1968 by the Dutch government and became fully privatised in 1999. As an institute, CITO has been responsible for developing and constructing the national examinations of Dutch secondary education, which include the national reading examinations for

foreign languages. More information on the specific context of Dutch education will be given in chapter 6.

CITO offers any person interested in language testing free access to the so-called *Toetswijzer* (Toetswijzer.kennisnet.nl) and registered but free access to a digital newsletter. *Toetswijzer* is a practical and relevant site for teachers, examination coordinators, learners, and parents on a variety of issues related to language testing. Subjects are dealt with as varied as test preapartion, the latest info on the national examinations, quizzes and games, tips to reduce examination stress, or information on language testing notions and the latest developments in the field of language testing. Sites like these may help to set up collaborative dialogue and highlight and discuss the social and ethical responsibilities of language testers, test users and test takers.

A final characteristic of critical language testing we would like to discuss is the increasing recognition of the rights of test takers.

Rights of test takers

The test taker, let alone his/her rights, was not a major concern in traditional language testing. The test taker was seen as the provider of data used to compute the psychometric traits of a test. The knowledge and skills the test taker was expected to exhibit were determined by those who wrote the tests, leaving the test taker no space but to comply. Yet, the stakes of test can be very high indeed, as is e.g. the case of English tests in Asian countries, the new European countries, or areas such as Hong Kong, where mastery of the English language has become a decisive factor for future success and prosperity. Moreover, control over the knowledge and skills the test takers are expected to exhibit is in the hands of or determined by a school, education system government or industry. Hanson (1993) observes that:

> *In nearly all cases test givers are organizations, while test takers are individuals. Test-giving agencies use tests for the purpose of making decisions or taking actions with reference to test takers- if they are to pass a course, receive a certificate, be admitted to college, receive a fellowship, get a job or promotion. That, together with the fact that organizations are more powerful than individuals, means that the testing situation nearly always places test givers in a position of power over test takers* (1993:19).

Critical language testers explicitly aim at improving test takers' rights, next to their preference of collaborative dialogue and emphasis on the responsibilities of the test taker. Shohamy (2001) sums up some solutions offered by critical language testing.

> *With regard to consent, a test taker should be given the right to be tested and the right to refuse to be tested. There should also be honesty with regard to the purpose of the test, its practice and methods. Next, a test taker should be granted the possibility of being assessed by an alternative method other than the traditional 'test-only' system. Such information can be used as counter evidence against decisions based on tests only. In addition, as was argued in chapter 10 [The responsibilities of language testers], there is a need for sharing the power of tests by training the public in testing methods, in the testing process and in the rights of test takers. Testing cannot remain a field that belongs only to testers but rather test takers and the public at large need to be part of the discussion*
> (2001:158).

Collaborative dialogue, mutual responsibility of test giver and test taker, and the rights of the test taker are also prominent characteristics of a challenging perspective on language testing called dynamic language testing. In the next section we will go

into the second approach that lent its name to the trend we labelled as critical-dynamic.

Dynamic assessment

As the opposite of static, the word dynamic holds the promise of energetic change and progress. *Dynamic assessment* is a direction in the field of language testing that calls for such a change. Its origins are closely linked with Vygotsky's sociocultural theory.

Luria (1961:7) was the first to contrast 'statistical' with 'dynamic' approaches of assessment. He felt that statistical assessment by way of a solo performance on a test, even though it was firmly grounded in psychometric principles, did not present a complete picture of what an individual was capable of. With this view, he followed Vygotsky (1956), who had shown that a child's actual development was not a valid indicator of his/her potential development. As an example, Vygotsky had offered the cases of two seven-year-olds, who both managed to solve tasks accessible to children of their age. Yet, with the help of leading questions, examples, and demonstrations, one of the children was able to solve tasks that corresponded with a mental age of nine, whereas the other child was only able to solve test items that were half a year above its level of actual development. It led to Vygotsky's explorations and demarcations of the ZPD, which we discussed as one of the main themes of sociocultural theory in chapter 3. Dynamic assessment is closely linked with the notion of the ZPD.

Luria suggested dynamic assessment (DA) as an alternative to statistical assessment. DA assesses an individual's performance with assistance from someone else as well as the extent to which that person can benefit from this assistance, not only in completing the test or task, but also in view of future performance in other tests or tasks. Dynamic assessment sharply contrasts with traditional approaches of non-dynamic assessment, where:

> *the examiner presents items, either one at a time or all at once, and each examinee is asked to respond to these items successively, without feedback or intervention of any kind. At some point in time after the administration of the test is over, each examinee typically receives the only feedback he or she will get: a report on a score or set of scores. By that time, the examinee is studying for one or more future tests.*
> (Sternberg & Grigorenko 2002: vii)

In traditional assessment, feedback tends to be limited and interventions in the course of the assessment are rejected. Because any intervention during the conduct of an assessment is likely to change a person's performance, it threatens the psychometric principles on which traditional assessment is generally based, particularly test reliability (See Lidz, 1991, Haywood, Brown and Wingenfeld, 1990). Drawing on Valsiner & Van der Veer (1993), Lantolf & Thorne define DA as 'a future-in-the-making model where assessment and instruction are dialectically integrated as the means to move toward an always emergent (i.e. dynamic) future rather than a fixed end-point' (2006: 330).

There are two main methodological differences between DA and traditional forms of assessment. First, whereas traditional assessment focuses on the outcome of past development and supposedly matured abilities, DA highlights future development and the test taker's maturing mental functions. Second, in traditional assessment the test giver generally adopts a neutral or disinterested stance and does not intervene in the course of the assessment to minimize measurement error. In DA, the tester reassuringly intervenes and offers mediated assistance in the assessment process, e.g. by way of reflective questions, examples or demonstrations.

Lantolf & Thorne (2006) discuss two approaches of DA, which are distinguished from one another by the way in which the tester intervenes in the assessment process. The two approaches are identified as *interventionist* and *interactionist*.

In an *interventionist* approach of DA, the feedback is quantified and standardised. This is e.g. the case with a 'pre-test – intervention – post-test' design, in which the tester attempts to control for his interventions (Budoff, 1968). Another example of an interventionist approach of DA is offered by Guthke, Heinrich & Caruso (1986), who use a hierarchy of standardised hints for the tester to select from. An advantage of standardised feedback is that it can be computerised. However, interventionist approaches with their quantitative feedback do not necessarily address a test taker's ZPD. Vygotsky had already stated that 'we must not measure the child, we must interpret the child' (1998:204). That is why Minick (1987:127) argues against quantitative and in favour of qualitative approaches to DA. His argument is based on an understanding of the ZPD as 'a means of gaining insight into the kinds of psychological processes that the child might be capable of in the next or proximal phase of development and a means of identifying the kinds of instruction, or assistance that will be required if the child is to realize these potentials.' Such a non-psychometric perspective underlies an *interactionist* approach to DA.

The works of Feuerstein et al. on the so-called *mediated learning experience* (MLE) are examples of interactionist DA (see Feuerstein, Rand, Hoffman & Miller, 1980; Feuerstein, Falik, Rand, & Feuerstein, 2003). In MLE, the traditional tester/testee roles are abandoned in favour of a teacher/learner relationship in which both are working towards the ultimate success of the learner. Feuerstein describes MLE as a process through which environmental stimuli are filtered through some other person, usually an adult mediator, who selects, frames, modifies, and imposes order to ensure that 'the relations between certain stimuli will be experienced in a certain way' (Feuerstein, Rand & Rynders, 1988:56). Feuerstein, Rand and Rynders outline eleven components of MLE, of which Lantolf & Thorne (2006) select three that are directly relevant to a discussion of DA: *intentionality, reciprocity,* and *transcendence*. We will briefly discuss the three components.

Intentionality refers to an adult's deliberate effort to mediate for the child external stimuli such as the world, an object in it, or an activity. For Feuerstein, intentionality distinguishes the MLE from the haphazard, incidental nature of traditional instruction that generally fails to take account of an individual's ZPD.

A second component of the MLE relevant to DA is reciprocity, which indicates that mediator/learner interaction is necessarily intertwined. In an MLE session, the learner is not a passive recipient of knowledge, but an active co-constructor of it. The final component is called transcendence. The notion relates to the ultimate goal of the MLE, which is to have the learner move from the 'here-and-now' of an activity to future and possibly more complex actions and activities. MLE typically proceeds from an initial training phase on a particular issue to the tackling of 'a series of tasks that represent progressively more complex modifications of the original training task.' (Hoffman, 1979:92).

The notion of transcendence runs counter to the often-voiced concerns with 'teaching to the test', in which the purpose, type or content of a test dominates the teaching/learning process.

An immediate pedagogical issue Feuerstein had to resolve is how to structure an MLE. By including tasks of varied level and complexity, Feuerstein et al. structure the MLE in such a way that learners are required to make the same kinds of adaptations as they are expected to in daily life.

The nature of the MLE and DA has informed a series of studies that focus on L2 learners or bilinguals (Peña & Gillam, 2000; Kozulin & Garb, 2002; Antón, 2003).

We would like to finish our review of DA with a discussion of the relationship between DA and formative assessment. We feel this discussion is of particular relevance to the research focus of *Testing for Autonomy.*

DA and formative assessment

In our definitions of assessment and evaluation, we already referred to the difference between formative and summative tests. Summative assessment and its corresponding evaluation take place at the end of a unit of study, a course, or any programme used for purposes of accountability, admission decisions, promotion, or selection. (See d'Anglejan, Harley & Shapson, 1990; Torrance & Pryor, 1998). Formative assessment takes place during a course of study and is generally used to gather:

> *information which will inform teachers and students about the degree of success of their respective efforts in the classroom. It allows teachers to diagnose students' strengths and weaknesses in relation to specific curricular objectives and thus guides them in organizing and structuring instructional material.*
> (d'Anglejan, Harley & Shapson, 1990:107)

Ellis (2003:12) identifies two general types of formative assessment and evaluation (FA), i.e. *planned* and *incidental*. Planned FA involves direct testing of learners and scales the present state of their knowledge or abilities during a course of study. Incidental FA takes place in the course of the instructional conversations that arise between teachers and learners in regular classroom pedagogical activity (Ellis: 2003: 314). Incidental FA may be *internal* or *external*. Internal incidental FA occurs 'through teacher questioning and probing' and provides learners with feedback on their performance as it is unfolding. It contributes directly to the accomplishment of a certain task, and may at the same time indirectly contribute to language development, presumably because it helps learners to understand what is expected of them and allows them to compare their present achievements with the desired performance (Ellis, 2003: 315). External incidental FA, on the other hand, refers to the teacher and learners reflecting on learner performance either during or following completion of an activity.

Vandergrift & Bélanger (1998:572) feel that FA has the added feature of motivating learners by providing them with feedback about what they can do already and what needs to be improved. Rea-Dickins & Gardner (2000:229-30) conclude that FA serves teachers in four different ways: it helps them plan and manage their teaching; it provides evidence of student learning; it indexes the extent to which they and their students have attained what has been described in the curriculum; and it provides them with evidence for evaluating their own teaching. However, there is one main criticism of FA, despite its obvious benefits. Rea-Dickins & Gardner (2000:31) point out its problematic validity and appropriateness, caused by the fact that FA generally involves informal procedures and is typically unsystematic. This means that FA might result in erroneous evaluation, with the result that learners might either be offered inappropriate instruction or no instruction at all when it is in fact required. A critique of the focus of research on FA is worth mentioning as well. Dann (2002:142) puts forward that the FA studies have primarily investigated 'the ways in which teachers have tried to inform their own practice so that pupils' needs are more specifically met'. Much less attention has been paid to 'the ways in which pupils participate in this process'. Poehner & Lantolf (2005) feel that a DA approach to formative classroom assessment has the potential to counter the criticisms of FA and FA research. According to Lantolf & Thorne (2006: 349), DA approaches are systematic and focus on 'mediation that is attuned to the learner's or group's responsiveness to assistance and, at the same time, promotes the very development it seeks to assess in the first place'.

Lantolf & Thorne do acknowledge, however, that more research is needed into the nature of dynamic assessment. Very much the same can be said of formative

assessment and its corresponding evaluation, albeit planned, external incidental, or internal incidental in Ellis's terminology.

Our discussion of the critical-dynamic trend in language testing closes off the four perspectives on language assessment and evaluation we singled out. Below, we will briefly look back on the perspectives we discuss and introduce the third section of this chapter on professional standards in second and foreign language assessment and evaluation.

Four perspectives in retrospect

Our discussion of the four different perspectives on language testing has shown how historical, social and cultural contexts have determined views on language assessment and evaluation. In the pre-scientific trend, assessors primarily rely on intuition and common sense and assess and evaluate learners in what they feel to be the best possible ways. An interest in behaviourism, structuralism and psychometrics gradually turned language assessment and evaluation into positivist science. Language testers favouring the psychometric-structuralist trend go for accuracy and high levels of reliability of their predominantly discrete-point tests. This leads to a community of language testing professionals, which is progressively influenced by debates on the nature of language ability inspired by sociolinguistics. It results in an integrative-sociolinguistic trend in language testing, with its interest in communicative ability, integrative tests and types of validity as essential measurement qualities. In the meantime, societies continue to change rapidly and often radically. It leads to points of view we defined as post-modern. In our post-modern age, conventions are critically challenged, innovation and change embraced, ambiguity is welcomed, and diversity emphasised. Postmodernism has gradually affected the field of language testing. Spolsky's 1999 agenda for language testers, critical language testing and dynamic assessment are examples of post-modern influence in a more recent trend we labelled as critical-dynamic.

Most of the perspectives on language testing we discussed welcome professional standards. In that sense, the language testing community attempts to be eclectic and tries to merge their belief in essential measurement qualities with new insights in language assessment and evaluation. In the third section of this chapter, we will dive more deeply into what these professional standards entail. We feel such information may help to bridge gaps between teachers as intuitive examiners and language testing professionals. It may also provide the topical knowledge required for informed discourse on second or foreign language assessment and evaluation between testers and teachers.

5.4 Professional standards in assessment and evaluation

Professional standards in language testing relate to a number of important domains. We have selected three of these subject areas. We will first highlight the importance of being explicit about the purpose of a test. Test purpose determines what type of test is going to be used. We will subsequently discuss a number of test types and arrive at a general classification useful to this study. Next, we will discuss the essential measurement qualities of reliability and validity in some detail. Third, we will deal with Bachman & Palmer's (1996) model of test usefulness, which is related to their model of communicative competence we have referred to before. Bachman & Palmer define the use of a test that attempts to measure communicative language ability by six qualities that complement one another. The six components are reliability, construct validity, authenticity, interactiveness, impact and practicality.

5.4.1 Test purpose and test types

We have earlier mentioned that Spolsky (1999) stresses the importance of analysing test purposes and the construction of appropriate test designs. The quality of a language test can only be properly judged if its purpose is taken into account.

Tests can be used for a variety of purposes. They can be used to determine whether a candidate is admitted to a course of study (*entrance tests*), to place students in a group or class at a level appropriate to their degree of knowledge or ability (*placement tests*), or to discover a learner's specific strengths and weaknesses in view of decisions regarding future training, learning or teaching (*diagnostic tests*).

Another distinction of tests in terms of test purpose is the one between *achievement* and *proficiency tests.*

Achievement tests attempt to measure knowledge, skills or microskills that relate to a particular course of instruction or set of materials. Achievements tests are often teacher-made, because they explicitly link to a curricular teaching and learning process. Proficiency tests do not necessarily relate to a given curriculum or particular course of study. They focus on a future situation of language use and aim to measure performance-related knowledge and skills. This future situation of 'real-life' language use is called the criterion. The criterion is the domain of behaviour a test aims to assess, or, more specifically, an aspect of performance which is evaluated in scoring, such as fluency, accuracy etc. (McNamara, 2000:132). Large-scale proficiency tests are generally designed by measurement professionals, profoundly schooled in psychometrics and research methodology.

The relationship between a test and its criterion is a key notion in communicative language testing. Evaluation of an assessment procedure involves making inferences from the criterion on which the assessment procedure was based. We will return to the test-criterion relationship in more detail when we discuss the essential measurement quality of validity. If tests are meant to sample criterion behaviour they are called *criterion-referenced tests*. In the case of criterion-referenced measurement, a candidate's performance is interpreted in relation to predetermined criteria. These criteria can be verbal descriptions of a satisfactory performance at a given level. Emphasis is on the attainment of objectives rather than on an individual test taker's scores in relation to their ranking within a larger group of test takers. A candidate's performance is assessed and evaluated against descriptors that enable classification. McNamara (2000:132) defines a criterion as the domain of behaviour relevant to test design or an aspect of performance which is evaluated in scoring, e.g. fluency, accuracy etc.

There are two subtypes of criterion-related validity, i.e. *predictive validity* and *concurrent validity*. Predictive validity is an indication of how well a test predicts future performance in the knowledge or skill that is being tested. Concurrent validity is not about prediction, but validates the present ability of the test taker. If the scores of a test correlate highly with a recognized external criterion measuring the same area of knowledge or ability, the test is said to have concurrent validity

The measurement assumptions of *norm-referenced tests* differ from the ones of criterion-referenced tests. In the case of norm-referenced measurement, a candidate's scores are interpreted with reference to the performance of a given group, consisting of people comparable to the individual(s) taking the test. Norm-referenced tests focus on ranking individuals relative to a norm group or to each other.

Proficiency tests have over the years developed into communicative tests, which attempt to assess aspects of communicative language ability. What, then, characterises a communicative test?

Fulcher (2000) mentions three primary aspects of a communicative test. First, it should involve *performance*. The test taker performs an action, activity or task that relates as closely as possible to the criterion behaviour in a real-life situation. A test taker might for instance be asked to play the role of a foreign visitor who has lost

his/her way in an attempt to simulate, as authentically as one can, the real behaviour in a future situation. This brings us to the second aspect of a communicative test mentioned by Fulcher. A communicative test should be *authentic* in the sense that the test taker must be able to recognise the communicative purpose of a test task in order to respond appropriately. Besides, the input of the test task should not be simplified, and the tasks embedded in a real situational context. Third, communicative tests are scored on *real-life outcomes*, i.e. success on a test task is determined by the test taker achieving a satisfactory outcome. Our discussion of test purpose and test types takes us to a useful general distinction between tests made by Baker (1989).

Baker distinguishes between *system-referenced tests* and *performance-referenced tests*. System-referenced tests assess knowledge of language as a system. Baker states that 'Their aim is to provide information about language proficiency in a general sense without reference to any particular use or situation' (1989:10). In contrast, performance-referenced tests attempt to elicit information about the ability to use the language in a particular context, and are therefore directed at assessing a particular performance. R.Ellis mentions that 'system-referenced tests are more construct-oriented, drawing on some explicit theory of language proficiency, performance-oriented tests more content-oriented, drawing on a work-sample approach to design' (2003:284).

Both system-referenced tests and performance-referenced tests can be more or less direct or indirect. The adjectives refer to the important relationship between test performance and the criterion performance. Direct tests directly sample the criterion performance, with the aim to elicit a contextualised sample of the test taker's use of language. They are holistic in nature and generally need an external rating procedure to obtain the required measure of proficiency. Indirect tests are less contextualised, and are based on an analysis of the criterion performance, with the intention to obtain measures of particular knowledge or skills components that are considered essential for successful outcomes. They typically seek to assess proficiency by means of specific linguistic measures, which are directly obtained from the test itself, as in a Cloze test. Ellis acknowledges that 'No test (except, possibly, one based on an observation of testees performing some real-world task) corresponds exactly to the criterion it seeks to measure' (2003:284). Even though distinctions between direct, indirect, system-referenced and performance referenced may not always be absolute in a given test, the following table is useful in categorising the distinctions made by Baker.

	Direct (holistic)	Indirect (analytic)
System-referenced	Traditional tests of language ability: -free composition -oral interview	Discrete-item tests of linguistic knowledge: -multiple-choice grammar or vocabulary tests -elicited imitation of specific linguistic features -error-identification tests
	Information transfer tests: -information-gap -opinion-gap -reasoning gap	Integrative tests: -Cloze -dictation
Performance-tests	Specific purpose tests: -based on observing real-world tasks -simulations of real-world tasks	Tests that seek to measure referenced aspects of communicative proficiency discretely: -tests of specific academic sub-skills, e.g. the ability to cite from a published work -tests of the ability to perform specific functions or strategies, e.g. the ability to write a definition of a technical term

Table 5.1: Types of language assessment (in R.Ellis, 2003:285)

Direct or indirect system-referenced or performance referenced test tasks, may be part of a host of alternative assessment approaches, procedures, and test formats, such as portfolio tests, take-home tests, computer-adaptive tests, individual and group project tests and approaches geared at self-assessment and self-evaluation. We would like to end our discussion of system- and performance-referenced tests with McNamara's (1996) argument that communicative tests need to be both system- and performance-oriented.

Specification of the purpose of a test, the selection of an appropriate test type, and the objective to be as explicit as one can about the test-criterion relationship have considerably added to professional standards in language testing. It brings us to a second domain in language testing we wish to highlight.

5.4.2 Essential measurement qualities

A second concern that has led to professionalism and expertise in language testing has been the exploration of essential measurement qualities. Reliability and validity are generally considered the two notions that helped to set professional standards for effective tests of a second or foreign language. We will first discuss reliability and then move on to a more elaborate discussion of validity.

Reliability

Reliability is the consistency or stability of measurement of individuals by an assessment procedure. Reliable tests are tests that produce consistent and stable scores, i.e. the scores are not affected by factors such as ambiguous instructions, the number of test items on particular knowledge or a particular ability that is tested, the lack of criteria for scoring and grading, or the temporal or spatial conditions of the test.

The more reliable a test is, the less random error it contains. In classical true score measurement theory, an observed score on a test is made up of two components: a *true score* reflecting the person's knowledge or ability, and an *error*

score, which reflects the influence of chance factors that are unrelated to the targeted learning. Chance factors, such as ambiguous or unclear instructions, produce error that is random or *unsystematic*. Investigation of reliability usually concerns both the consistency of the scores assigned to a test item (*score reliability*) and the consistency with which various raters assign the same or similar scores to a test item (*rater reliability)*.

Next to unsystematic error, there may be factors in a test which regularly affect the performance of some individuals, causing systematic error or test bias. A test or item is considered to be biased if one particular section of the candidate population is advantaged or disadvantaged by some feature irrelevant to what is being measured. Sources of bias may be connected with gender, age, culture etc. Systematic error is generally related to the concept of validity. Test bias may be hidden to the extent that tests are reliable, but nevertheless invalid.

A criticism sometimes heard is that measurement professionals put most of their emphasis on reliability issues, often at the expense of the validity of assessments. The emphasis on reliability issues is generally justified with the claim that assessment procedures first of all have to be reliable, before its validity can be considered. Nevertheless, it is a test giver's responsibility to construct tests that are as dependable as they can possibly be. How acceptable levels of reliability are ensured can be illustrated by having a closer look at stages of professional test construction.

Stages in professional test construction

In the case of professional test construction, the process of test design is typically quite explicit and open to mediation and adaptation before a final version of a test is marketed. An example of open and explicitly formulated stages of test construction is provided by CITO[1]. They distinguish eight clearly defined steps in the process of test construction.

1. *Specification of test purpose and objectives. The first stage of test construction is the operationalisation of the knowledge or skills the test is expected to measure and specification of the ways in which the test is going to be used. The procedure allows for a consistent sequence in the process of test construction that does justice to four requirements of language tests: reliability, validity, acceptability and transparency.*
2. *Test specification. The second stage involves a number of questions related to the domain of the test, i.e. the defined area of content and/or ability that is to be tested by a specific task or component of the test. The specifications give details of the design, content, level, types of items and tasks, target population etc.*
3. *Item construction. Once the function of the test is clear and the specifications are given, the items can be constructed. The items and tasks that a test contains are not selected at random. They are closely related to the objectives of the test. Test objectives are often related to goals formulated at a national level, e.g. objectives formulated as competencies in a national curriculum.*
4. *Test administration. Professional tests are always piloted before a final version is made available. First versions are tried out and revised, certain items may be removed, others reformulated, after which the test may be piloted again. Conditions of test administration and the rating procedures are standardized, i.e. it is ensured that assessors adhere to agreed administration and rating procedures.*
5. *Item evaluation. Psychometric measurements are used to determine the quality of a pilot test. Common methods are the determination of facility- and discrimination indexes. A facility index indicates the proportion of correct responses to an item on a scale of 0 to 1. Sometimes it is expressed as a percentage. Popular names for the facility index are 'facility value' or 'p-*

value'. Discrimination indexes indicate in how far an item discriminates between test takers with high test scores and those with low ones.

6. *Test design. The final design of the test has to conform with three requirements, i.e. (1) psychometric demands related to the adequacy of measurement ; (2) educational demands related to test content, such as the distribution of items of specified content categories or the particular level of difficulty required; (3) practical demands, i.e. requirements related to resources such as available time, money, number of computers, number of raters etc.*
7. *Domain of reference. Test scores on a test become meaningful when they are interpreted in relation to some norm (norm-referenced measurement) or criteria (criterion-referenced measurement).*
8. *Test manual. As a final stage, a test manual is prepared, in which the test quality is justified elaborately. The test manual provides detailed information on test purpose, the origin of the constructs the test attempts to measure and the test-criterion relationship. In addition, the manual provides information on the way in which administration and rating procedures are standardised, followed by information about the specific standards of performance used in the test. Finally, information is given how appropriate levels of validity and reliability have been maintained*

(Source CITO http://toetswijzer.kennisnet.nl, translated from Dutch)

The eight stages illustrate how professional test constructors attempt to achieve not only acceptable levels of reliability, but also do justice to the requirements of acceptability and validity by being transparent, precise, and procedural. We assume that practising teachers lack the time and perhaps the expertise to follow the steps in test construction as meticulously as testing professional do. Nevertheless, we feel it is essential to address both score and rater reliability in classroom-based assessment and evaluation. In view of the often high stakes of teachers' informal tests, we feel that learners are entitled to dependable assessment procedures. What can non-specialists do to ensure tests that are acceptably reliable?

Teachers can generally make tests more reliable if the test is constructed, administered, scored and graded in close cooperation with others. Attention should be paid to factors such as the clarity of instructions, over- or under representation of certain test items, use of test items that are formulated independent of one another, comparable test conditions of time and place when a test is administered more than once, acceptable differences in the levels of the learners sitting the test, and unambiguous criteria for scoring and grading.

However, reliable tests are not necessarily valid tests. In the next paragraphs we will probe more deeply into the important but complex test quality of validity.

Validity

So far, we have either directly or indirectly referred to the crucial measurement quality of validity. A bare definition of test validity is deceptively simple. A test is supposed to be valid if it measures what it is expected to measure. More often than not, such a basic interpretation of test validity only leads to surface acceptability of the test by those involved in its development or use. On the face of it, the assessment procedure seems valid. This type of validity is therefore referred to as *face validity*. Yet, the notion of validity is far more complex than basic interpretations suggest. Several types of validity have been identified. In this section, we will explore test validity in more detail.

We have defined testing as an assessment procedure meant to collect data that provides evidence of learning. Whether the response to test items is considered as evidence is a matter of judgment and interpretation. It depends on the abstract inferences a test user makes from the empirical data reliably gathered by a test or any other assessment procedure. Test users -teachers, learners, school managers,

politicians, and parents alike- implicitly or explicitly draw conclusions on the basis of test performance. These conclusions should be defensible and fair. McNamara (2000) defines validity as the extent to which particular test scores enable inferences that are appropriate, meaningful and useful in view of the purpose(s) of that test. The appropriateness, meaningfulness, and usefulness of these interpretations have to be proven beyond reasonable doubt. The legal overtone here is not incidental. There are indeed similarities with legal cases. Sometimes legal evidence can be so overwhelming that only one conclusion can be drawn by the judge and/or jury. However, more often than not, the interplay of hard facts, reasoning and interpretation is less straightforward.

> *In the case of both legal and assessment settings, the focus of investigation is on the procedures used. If the procedures are faulty, then conclusions about particular individuals are likely to be unsound. The scrutiny of such procedures will involve both reasoning and examination of the facts. In the legal case, the reasoning may involve legal argumentation, and appeals to the common sense, insight and human understanding of the jury members, as well as careful examination of the evidence. Test validation similarly involves thinking about the logic of the test, particularly about its design and intentions, and also involves looking at empirical evidence – the hard facts – emerging from the data from test trials or operational administrations. If no validation procedures are available, there is potential for unfairness and injustice. This potential is significant in proportion to what is at stake.*
> (McNamara, 2000: 48)

Similar to some legal cases, finding valid proof of learning beyond reasonable doubt is no easy matter. Convincing evidence of second and foreign language learning is particularly difficult because learning in general is a mental trait that is hard to observe or measure directly. McNamara (2000:50) singles out three problem areas that tend to threaten the meaningfulness, interpretability, and fairness of the scores or ratings of a given assessment and evaluation procedure. Threats to validity might be caused by test *content* (what the test contains), test *method* (the way in which the test taker is expected to engage with the materials and tasks in the test), and test *construct* (the underlying ability the test attempts to measure).

The contents of a test have to adequately sample and be representative of the *domain* of the test, that is the area of knowledge or skill or the set of tasks constituting criterion performance, and which is the target of the test. If learners are given an informal test on what they have learned in the past two units of a foreign language coursebook, that test is said to have effective *content-related validity*, or *content validity* if the knowledge and skills of the units are evenly and adequately represented in the test. If the course units pay a lot of attention to oral skills, and the resulting paper-and-pencil test primarily measures knowledge of grammar and vocabulary, the test is said to have low content validity. This is because no valid inferences can be made regarding the oral content and skills that have been part of the course. If the contents of a course unit comprise both system-related and performance-related knowledge and/or skills, as course units of communicative courses often do, then the test has to represent and adequate sample of each. Next to inadequate sampling of content, there are other threats that jeopardise test validity.

The implications of the choice for a particular test method or test construct might also endanger the meaningfulness of the interpretations that are made of the results on that test. The test method is the manner in which a test taker is asked to engage with the materials, items or tasks in a test, and how their response will subsequently be scored. The test method should adequately elicit the test construct, that is the underlying ability or trait the test aims to assess. Variations in the ways in which a candidate engages with the materials might affect the scores on a test. Examples of such variations are allowing candidates to take notes while listening, permitting them

to listen to a text more than once, exposing them to longer or shorter chunks of oral or written text, or the effect of the use of dictionaries on the final scores of a test. The latter example may cause low *construct validity* if the objective of the test is to sample the test taker's productive mastery of vocabulary. Even when the test contents and test method used to elicit response, other aspect might endanger the construct validity of a test. McNamara (2000:52) observes that 'rating procedures introduce a host of variables in the assessment'. Such variables are examples of factors that are irrelevant to the aspect of ability being measured (*construct irrelevant variance*) or instances of factors that require too little of a candidate (*construct under-representation*).

The notion of construct validity is central to this investigation. After all, we aim to explore how the two constructs of learner autonomy and communicative competence are reflected in the teacher respondents' assessment and evaluation practices. That is why we will explore the notion of construct validity a little more in the next section.

Construct validity

Messick (1989, 1994, 1996) is generally seen as one of the most influential theorists on validity. He did seminal work in developing validity theory. Messick (1989) proposed that construct validity embraces almost all forms of validity evidence and that it should be regarded as a unitary concept. He felt that all of the traditional forms and types of validity are varieties of evidence that supplement one another. It resulted in a multidimensional conception of validity as 'an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment'(1989:13). Messick identified six sources of validity evidence, which together determine the validity of an assessment procedure.

> *We can look at the content of a test in relation to the content of the domain of reference. We can probe the ways in which individuals respond to the items or tasks. We can examine relationships among responses to the tasks, items or parts of the test, that is, the internal structure of test responses. We can survey relationships of the test scores with other measures or background variables, that is, the test's external structure. We can investigate differences in these test processes and structures over time, across groups and settings, and in response to experimental interventions – such as instructional or therapeutic treatment and manipulation of content, task requirements, or motivational conditions. Finally, we can trace the social consequences of interpreting and using test scores in particular ways, scrutinizing not only the intended outcomes but also unintended side effects*
> (1989:16)

According to Messick, validity evidence is made up of half a dozen of interrelated components. Messick first refers to an analysis of content relevance and content representation. Next, he refers to correlational studies of structure and variance components, both internal to the test with a study of item or task consistencies, and external to the assessment procedure by relating the test scores to other variables. Then he deals with the examination of the processes underlying item responses and task performance, with specific attention to longitudinal, interventionist, and manipulative aspects. Finally, he includes a consequential component in his unified conception of construct validity.

Messick's paper on validity has been inspirational and challenging to the theorist, but tends to be mind-boggling and complex to the practitioner. For that reason, Taylor & Nolen (1996) situate Messick's components of validity evidence in the context of classroom teachers' decision-making. They question the utility of traditional conceptualisations of validity and reliability, developed in the context of large scale, external testing, and the psychology of individual differences, for the context of the classroom. Taylor & Nolen reduce Messick's six components to five

dimensions of validity evidence, and discuss these in the context of formal education. Drawing on Taylor & Nolen, we will briefly summarise the five dimensions and the ways in which the authors feel the reduced sources of validity evidence can be put to good use by teachers.

Validity Dimension 1: Looking at the content of the assessment in relation to the content of the domain of reference.

Taylor & Nolen feel that a study of content relevance and content representation should be preceded by the teachers thinking clearly about their disciplines, understanding both the substantive structure (critical knowledge and concepts) and the syntactic structure (essential processes) of the disciplines they teach (Schwab, 1978). Once teachers have clearly conceptualised the disciplines they teach, they must know how to ascertain the degree to which the types of assessment tasks used in the classroom are representative of the range and relative importance of the concepts, skills, and thinking characteristic of subject disciplines. Besides, teachers should also examine in how far their ways of scoring and grading reflect the targeted learning. Such activities on the part of the teachers are important, because without a clear picture of what is to be accomplished in a course or subject area, teachers cannot adequately assess whether their assessments (selected or self-developed) are valid.

Validity Dimension 2: Probing the ways in which individuals respond to the items or tasks and examining the relationships among responses to the tasks and items.

Taylor & Nolen first acknowledge that teachers do not often have the luxury of ‘item tryouts’ when developing their assessments. They nevertheless discuss four teacher actions and activities that may help them to collect evidence related to validity dimension 2. First of all, teachers can examine whether the test items and tasks actually elicits the learning aimed for. Secondly, they should be challenged to develop and use assessment strategies that will allow them to probe their students' thinking and reasoning processes. This is particularly important in view of the attention paid to learner autonomy and higher-order thinking skills (Lin & Mackay, 2004). Such probing does not only provide information about the validity of the assessments, but also provides for better pictures of students' learning. Thirdly, teachers must know how to look across students' responses to a variety of assessment tasks to determine whether certain response patterns support the use of the assessments. Teachers can e.g. be shown how to scrutinise student work qualitatively, looking for patterns in responses that reveal positive and negative information about the assessments. Finally, in order to probe examinee performance within and across different measures, teachers can learn to develop multiple measures of the same targeted learning. They may not only discover different ways to assess a given construct, but they may discover for themselves that particular types of assessment are more or less suited to certain learning targets.

Taylor & Nolen’s second dimension of validity evidence is in fact a combination of three of Messick’s original components, that is the so-called *substantive* component, the *structural* component and part of the analysis of *processes* component. After the content component, Messick (1989) refers to the substantive component, which ‘entails a veritable confrontation between judged content relevance and representativeness, on the one hand, and empirical response consistency, on the other’ (1989:42). Attention to the substantive component helps to rationally attune the scoring model to the nature of the construct. Then, Messick goes into the structural component of construct validity. This component is in fact a further specification of the substantive component, which seems to justify why Taylor & Nolen reduced the two components to one dimension. The structural component specifically concerns

investigations of the internal structure of a test. It is studied in how far the nature and dimensionality of the inter-item structure of a test reflects the nature and dimensionality of the construct domain. This means that scoring models should be rationally consistent with the nature and dimensionality of the construct domain. If a construct is multifaceted, the scores should be based on separate scales as well. This means that assessing a construct such as literacy, would typically result in a profile of scores for several categories. The next source of validity evidence Messick discusses is the processes component, which addresses analyses of processes, group differences and changes over time, responsiveness of scores to experimental treatment, and finally manipulation of conditions. Part of Messick's processes component is present in Taylor & Nolen's next validity dimension.

Validity Dimension 3: Investigating differences in assessment processes and structures over time, across groups and settings, in response to instructional interventions.

This validity dimension specifically focuses on examining the relationship between the instructional practices used and the assessments themselves. Examination of this dimension of validity can be obtained when teachers are asked to look carefully at the relationship between an instructional plan and the demands of an assessment. To investigate these validity issues, teachers must examine in how far their instructional practice has been adequate for their learners to exhibit the targeted learning, and in how far instruction has matched the needs of learners in different groups and settings. This validity dimension assumes that teachers can develop and use a host of instruction and assessment strategies.

Validity Dimension 4: Surveying relationships between assessments and other measures or background variables.

Teachers must ensure that their learners' performance and the scores resulting from the assessment are directly attributable to the targeted learning. Taylor & Nolen primarily mention that teachers should pay attention to factors irrelevant to the targeted learning such as assessment format, response mode, gender, or language of origin. They should also investigate the relevance of their scoring mechanisms in view of the learning they wish to elicit. In discussing this dimension, Taylor & Nolen do not go into the importance of comparing and contrasting (the results of) an assessment with other assessments, which is crucial in Messick's *external component* of construct validity (1989: 45).

Validity Dimension 5: Tracing the social consequences of interpreting and using test scores in particular ways, scrutinizing not only the intended outcomes, but also the unintended side effects.

Teachers must realise that the nature of the assessments, feedback, and grading can all influence student learning, students' self concepts and motivation (Butler & Nisan, 1986; Covington & Omelich, 1984). It may also affect their perceptions of the disciplines they are being taught. Awareness of washback and impact of assessment procedures seems a crucial component of validity evidence. Teachers can be asked to look at their methods of feedback (formative assessments) and determine whether they are likely to motivate learning or to stifle learning; to assess whether feedback will lead to improvement, be largely insubstantial (Sommers, 1982), or be perceived by students as too late to make a difference in their grades (Canady & Hotchkiss, 1989).

This particular component of validity has come to be known as *consequential validity*. It implies an investigation of any intended or unintended changes that may occur because of the use of a particular assessment procedure. Messick (1996) is explicit about how he sees the consequential component of construct validity.

> *The consequential aspect of construct validity includes evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short- and the long-term, especially those associated with bias in scoring and interpretation, with unfairness in test use, and with positive and negative washback effects on teaching and learning.* (1996:251)

Messick feels that 'consequential validity' or, worse still, 'washback validity', should not be viewed in isolation as a separate type of validity. He refers to it as a consequential component of validity evidence that is part of a unified conception of construct validity. This particular source of validity evidence primarily focuses on the learners themselves. Nevertheless, consequential validity is strongly linked with construct validity. Teacher beliefs and construct definitions are likely to impact *what* is tested, *how* this is done, *why* this is done in a particular assessment format. What teachers believe and how teachers perceive important constructs of their disciplines are likely to affect the interpretability of the results of an assessment procedure. Messick warns about any negative washback effects on teaching and learning that are caused by construct under-representation or construct-irrelevant variance.

> *The primary measurement concern with respect to adverse consequences is that negative washback, or, indeed any negative impact on individuals or groups should not derive from any test invalidity such as construct under-representation or construct-irrelevant variance … . That is, invalidly low scores should not occur because the assessment is missing something relevant to the focal construct that, if present, would have permitted the affected persons to display their competence. Moreover, invalidly low scores should not occur because the measurement contains something irrelevant that interferes with the affected person's demonstration of competence.* (Messick, 1996: 252)

Messick has been given the first and the last word in this section on construct validity, which meant to further specify the notion and unravel some of its complexities. Both reliability and validity are essential measurement concerns that cannot and should not be neglected in formative and summative classroom assessment and evaluation.

As a final instance of professional standards in second and foreign language assessment and evaluation, we will discuss a model of test usefulness developed by Bachman & Palmer. The model directly derives from the models of communicative competence we have referred to before (Bachman, 1990; Bachman & Palmer, 1996). The model, thus, combines what is known about the nature of communicative ability with what is known about language testing.

5.4.3 Bachman & Palmer's model of test usefulness

Successful tests are useful tests. Effective second or foreign language tests attempt to provide reliable and valid information on the degree of success of what has been taught and/or learned in some instructional programme or on what a candidate actually knows about or is able to do with the target language in a particular situation. Bachman & Palmer have explored the notion of test usefulness. Their exploration is based on the communicative model they developed, which we presented as an example of the communicative models that resulted from integrative-sociolinguistic interests in language testing in 5.2.3 of this chapter.

Central to Bachman & Palmer's conception of test usefulness is the notion of Target Language Use (TLU) domain, with its derivative TLU task. They define a target language use domain as 'a set of specific language use tasks that the test taker is likely to encounter outside of the test itself, and to which we want our

inferences about language ability to generalize'(1996:44). They distinguish two types of TLU-domains. *Real-life domains* consist of tasks used in real-life situations, where language is used for purpose of communication. An example of a real-life domain could be 'English for tourist information'. Within this domain, a variety of different settings can be identified, dependent e.g. whether a person gives tourist information on a particular area or whether (s)he wishes to get this information; whether that person gives or receives that information orally or in writing; whether it relates to information on places to stay or on activities to do etc. A specified TLU domain would contain all the foreign language and communicative behaviour needed to successfully communicate in a given situation. The other type of TLU-domains consists of situations in which language is used for the purpose of teaching and learning another language. Bachman & Palmer refer to this type as a *language instruction domain.* Thus, a TLU task is a task used within a specific TLU domain, either related to real-life or related to an instructional programme.

Bachman & Palmer argue that a useful test should invariably include tasks with the characteristics of TLU tasks.

> *Language use tasks can be thought of informally as constituting the elemental activities and situations of language use. That is, language use can be viewed as the performance of a set of interrelated language use tasks. A language test can be thought of as a procedure for eliciting instances of language use from which inferences can be made about an individual's language ability. It therefore follows that in order for such inferences to be made, a language test should consist of language use tasks*
> (1996:45)

Bachman and Palmer feel that the key to designing tests that are useful for their purposes is the inclusion of test tasks whose distinguishing characteristics correspond to those of TLU tasks. In the next section, we will present how the two theorists see the notion of test usefulness.

Components of test usefulness

Bachman and Palmer (1996:18) express the notion of test usefulness by way of the following equation:

> Test usefulness = reliability + construct validity + authenticity + interactiveness + impact + practicality

The usefulness of an assessment procedure is seen as the combined effect of six interrelated qualities. The authors stress that the separate qualities must be evaluated in terms of their united effect on the overall test usefulness. We see that reliability and construct validity, the two essential measurement qualities we discussed above, are part of the six qualities of test usefulness. The remaining qualities of authenticity, interactiveness, impact and practicality call for further explanation.

Authenticity is seen as 'the degree of correspondence of the characteristics of a given language test task to the features of a TLU task' (1996:23). Bachman & Palmer see task authenticity as an important test quality, because it relates the test task to the TLU domain to which the score interpretations are meant to generalise. Thus, authenticity is directly linked to construct validity.

Interactiveness is a test quality that initially conjures up a different connotation than the one intended by Bachman & Palmer. At first sight, one tends associate interactiveness with interaction, which is a general characteristic of communicative encounters. However, interactiveness in the authors' terms refers to the test takers' often complex and implicit mental operations when they accomplish a test item or

task. Interactiveness is defined as 'the extent and type of involvement of the test taker's individual characteristics in accomplishing a test task (1996:25). The individual characteristics most relevant to second or foreign language testing are the test taker's language ability (language knowledge and strategic competence, or metacognitive strategies), topical knowledge, and affective schemata. Topical knowledge can loosely be defined as subject knowledge that has been retained in long-term memory. More often than not, such knowledge has been determined and shaped within certain cultural contexts. Affective schemata 'can be thought of as the affective or emotional correlates of topical knowledge' (1996:65). The test taker's emotions and affective responses can either facilitate or hinder the accomplishment of a certain task.

Impact is a test quality we have referred to before. Bachman & Palmer define impact as the ways in which a test and/or the use of test scores 'impact on society and educational systems and upon individuals within those systems' (1996:29). They see washback as an instance of impact on individuals. In the case of formal education, tests may impact on the teachers as test givers and on the ways in which they instruct as well as on the learners and the ways in which they learn.

Finally, practicality is considered an important characteristic that determines test usefulness. It pertains to the ways in which the test will be implemented, rather than to its use. For tests to be practical, the available resources should at least match the required resources. Examples of resources are the number of people involved in test construction, administration, and scoring (human resources), material resources (space, equipment, and materials), and time (development time as well time for specific tasks, such as scoring and grading).

The six qualities of test usefulness defined by Bachman & Palmer have the potential to improve language testing in practice. It may help non-specialist practitioners to improve their assessment and evaluation practices. Bachman & Palmer give ten examples of projects in which the notion of test usefulness literally has been put to the test. More recently, Douglas (2000) used Bachman & Palmer's framework to analyse TLU situations and test task characteristics, with its typical approach of test rubric, input, the expected response, the interaction between input and response, and finally, the assessment criteria.

In the last two sections of this theoretical chapter, we will first summarise the issues and notions we have just discussed. After that, we will relate the chapters on learner autonomy and communicative competence to the present one on language assessment and evaluation and explain how the issues, notions and ideas of the three theoretical chapters are going to be used in the study.

5.5 Summary

In the first part of our chapter on second and foreign language assessment and evaluation, we first of all argued for clear and straightforward definitions of the notions of *testing*, *assessment* and *evaluation*. We observed that assessment and evaluation are often used interchangeably. When the notions of assessment and evaluation are used, it is not always clear whether a person refers to actual measurement procedures or to the inferences made from their results. We felt the need to distinguish between the two in the light of our investigations. We use assessment for any measurement of knowledge or skills at a given time and evaluation for the inferences made on the basis of an assessment procedure.

The second section dealt with four trends in the field of language assessment and evaluation. We observed that the trends do not simply represent clearly defined periods. One or more of the trends we discussed may still be prevalent in a given language testing situation. We subsequently discussed the pre-scientific, psychometric-structuralist, integrative-sociolinguistic and the critical-dynamic trends.

Assessment and evaluation in the pre-scientific trend is generally linked with the grammar-translation method, which we discussed in chapter 4. Assessment is generally in the hands of a single assessor, who tests for linguistic accuracy and/or assesses oral or written skills globally and holistically. We stressed, however, that there have been teachers in pre-scientific times who had ideas about teaching and testing that were quite different from the general characteristics we mentioned.

Then we discussed the psychometric-structuralist tradition in language testing. Interest in science, psychometrics and American structuralism led to professionalism in language testing. Influenced by people like Cronbach or Lado, psychometric-structuralist testers aimed to provide "objective" measures, using various statistical techniques to assure reliability and certain kinds of validity. They developed predominantly indirect tests of a variety of discrete items or skills.

The integrative-sociolinguistic trend derives its name from integrative tests and from theoretical advances in the field of sociolinguistics. With its focus on language use, the integrative-sociolinguistic trend led to models of communicative competence and the development of tests that were meant to assess communicative ability. The integrative-sociolinguistic trend is still prevalent in language testing.

Perhaps a little less operationalised are the assessment procedures advocated by language testers of the critical-dynamic trend. Philosophical postmodernism challenged language testers into adopting an agenda for language assessment and evaluation with issues that were critical of existing language testing practices and in favour of more dynamic conceptions of assessment and evaluation. The following concerns were on the agenda by the end of 1990s: (1) an interest in language testing ethics, (2) detailed attention to test purpose and design, (3) the need for test procedures to be open to inspection and feedback, (4) a recognition of professional standards, and finally (5) a preference for multiple methods of assessment and evaluation. We then went into critical language testing, with its challenges of the power dimensions of test, collaborative and dialogical approaches to assessment, and attention to the ethics of language testing, such as the rights of test takers and the adoption of a code of ethics.

Finally, we dealt with dynamic assessment, an approach that draws on Vygotsky's sociocultural theory. Whereas traditional assessment tends to measure past development and matured abilities, dynamic assessment highlights future development and the test taker's maturing mental functions. In dynamic assessment, the test giver may intervene in the course of the assessment and offer mediated assistance by asking reflective questions, by giving examples, or by demonstration. We highlighted the relationship between dynamic assessment and formative assessment and concluded that an interactionist approach to dynamic assessment best serves the goals of formative assessment and evaluation.

In the third section we singled out three subject domains of language testing research that have helped to establish professional standards in the field. Because the quality of a test is strongly linked with the purpose for which it is used, we first discussed assessment procedures in terms of test purpose and test types. We distinguished between entrance, placement and diagnostic tests. Then we defined achievement and proficiency tests, with the first type being related to an instructional programme. The importance of the relationship between a test and its criterion resulted in a discussion of the relevance of criterion-referenced tests, with their subtypes predictive and concurrent criterion-referenced tests. Drawing on McNamara (2000), we defined a test criterion as the domain of behaviour relevant to test design or an aspect of performance which is evaluated in scoring, such as fluency or accuracy. Because proficiency tests are in fact communicative tests, we defined communicative tests as assessment procedures that involve performance, are authentic, and are scored on the basis of real-life outcomes. Finally, we divided communicative tests into system-referenced or performance-referenced direct (holistic) or indirect (analytic) tests and presented this division in a table. We

concluded that the specification of the purpose of a test, the selection of an appropriate test type, and the objective to be as explicit as one can about the test-criterion relationship have considerably added to professional standards in language testing.

In a similar vein, we approached the essential measurement qualities of reliability and validity. They were the second domain of professional standards in language testing we singled out for discussion. We defined reliability as the consistency or stability of measurement of individuals by an assessment procedure, and discussed its related notions of true score, error score, unsystematic error, score reliability and rater reliability. As an illustration of how appropriate levels of reliability (and validity) are ensured, we presented eight stages in professional test construction. The procedure was referred to as transparent, precise and procedural, labels we will return to in the final two chapters of this study. We briefly mentioned what classroom teachers can do to make their assessments more reliable.

As the second important measurement quality we discussed validity, with its related notions of face validity, content-related validity, construct validity, construct irrelevant variance and construct under-representation. Referring to Messick (1989,1996) and Taylor & Nolen (1997), we then discussed the relevance to classroom-based assessment and evaluation of a multidimensional and unified conception of construct validity. We defined validity in Messick's terms as 'an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment'(1989:13).

As a final domain of professionalism in language testing, we discussed Bachman & Palmer's notion of test usefulness, which is firmly rooted in testing theory and based on their model of communicative language competence. Bachman & Palmer defined test usefulness as the sum of the six interrelated qualities of reliability, construct validity, authenticity, interactiveness, impact and practicality. We then discussed how the authors viewed the final four qualities of test usefulness.

Now we will look ahead at how the contents of the three theoretical chapters on learner autonomy, communicative competence and assessment and evaluation are going to be related to the data of our three respondent teachers *Joy*, the budding professional, *Mark*, the literary master, and *Pete*, the project man.

5.6 Looking ahead

Both learner autonomy and communicative competence have been discussed and specified as complex constructs that present challenging goals to second and foreign language learning. In this chapter we discussed the field of language assessment and evaluation. The contents of the three theoretical chapters will be used as theoretical domains with which the teacher data will be compared. The teacher data will be reported in four narrative chapters. Below, we will briefly refer what data we intend to generalize our findings to theory for each of the three constructs.

Our theoretical explorations of autonomy led to thirteen parameters that indicate both the limitations and the opportunities of learner autonomy as a pedagogical goal. We will generalise how the teachers define and/or attempt to implement learner autonomy in chapter 11. We will also highlight what the teachers consider to be effective foreign language teaching and learning, and what knowledge and skills they feel are required. Their views will then be compared to the learning theories we discussed in chapter 3. Next we will focus on how the teacher respondents view the relationship between learner autonomy and motivation. This data set will be interpreted in relation to the motivational theories discussed in the chapter on learner autonomy.

Our discussion of communicative competence resulted in an overview of approaches to the learning and teaching of another language. What the teachers consider to be effective teaching and learning, will be generalised to the methods we have highlighted and discussed in chapter 4. The teacher respondents were also asked to specify specific communicative knowledge and skills. These specifications allow for a generalization to the models of communicative competence we discussed. We will particularly compare and contrast how the teachers interpret communicative competence and in how far they focus on meaning and in how far on form.

The present chapters allows for comparisons with the ways in which the three respondent teachers constructed and used the tests they had selected for analysis and discussion. The assessment and evaluation practices associated with these tests will be compared and contrasted with the four trends in language testing. Then, we will relate the tests to the professional standards we discussed in chapter 5, that is, the importance of test purpose and test types, the attention paid to reliability and validity, and the aspects of test usefulness that were prominent in their ways of assessment and evaluation.

The approach outlined above, paves the way for the contents of the chapters to come. In the next chapter, we will provide details on foreign language education in the Netherlands and explore the curricular and didactic reform that partly determined the educational context in the year of data collection. Chapters 7 to10 are predominantly narrative chapters, which aim to bring the teacher respondents to life and do justice to their values and beliefs.

CHAPTER 6: A CONTEXT OF INNOVATION

6.1 Introduction

In the previous chapters we have explored three constructs relevant to foreign language education: *learner autonomy, communicative competence,* and *FL assessment and evaluation*. Before turning to the teacher data, information on the educational context may help to better understand the Dutch education system and the foci of learner autonomy and communicative competence in this investigation.

This chapter, therefore, concentrates on the structure of education in the Netherlands, curricular and didactic innovation, and the corresponding demands on our respondents and learners in the year of data collection. The three exploratory case studies were carried out in Dutch upper secondary education in a year of turbulent curricular and didactic reform.

In this chapter, we will first deal with the two fundamental legislative parameters of education in the Netherlands: *freedom of education* and *compulsory education*. Next, we will survey how secondary education in the Netherlands has been organised. Then we will focus on the three teacher qualifications in the Netherlands. In three subsequent sections, we will focus on English as a school subject in primary, basic and upper secondary education. Upper secondary education will be discussed in the light of the reform we have referred to as the *second phase*. The chapter will end with a brief summary.

6.2 Legislative parameters

Freedom of education and compulsory education are key features of formal education in the Netherlands. In this section we will briefly discuss the two.

Freedom of education is an important parameter of Dutch education. It concerns three essential liberties: the freedom to found schools (freedom of establishment), to organise the teaching in schools (freedom of organisation of teaching) and to determine the principles on which they are based (freedom of conviction). The liberties are guaranteed under article 23 of the Constitution. This means that people have the right to found schools and to provide teaching based on religious, ideological or educational beliefs. However, guidelines and attainment targets have been formulated for primary, secondary and higher education. The quality of education is monitored on a regular basis by the Department of Education, Culture and Science (Ministerie van Onderwijs, Cultuur en Wetenschappen) and the Education Inspectorate (Inspectie van het Onderwijs).

The Compulsory Education Act of 1969, and its review in 1985, laid down obligatory full-time time school attendance from the first school day of the month following a child's fifth birthday. However, most children attend primary school from the age of four. Children must attend school full-time for twelve school years and, in any event, until the end of the school year in which they turn 16. There are special programmes for children who drop out of the regular school system. Uninterrupted primary education lasts eight years, with learners aged from four to about twelve. Secondary school learners are aged from 12 to18, depending on the type of secondary education they attend. Uninterrupted secondary education lasts four, five or six years.

6.3 Education in the Netherlands

In this section we will have a closer look at how education has been organised in the Netherlands. At the end we will offer a graphic representation of compulsory education.

Children start primary school at the age of 4. The overall aim of primary education in the Netherlands is 'to promote the development of children's emotions, intellect and creativity and the acquisition of essential knowledge together with social, cultural and physical skills in an uninterrupted process of development. Teaching must reflect the fact that pupils are growing up in a multicultural society' (Eurybase, 2003: 4.6).

On leaving primary school, at the age of about twelve, the majority of learners move on to one of the three types of secondary education. A minority of learners, in particular those with learning or behavioural difficulties, go on to forms of special secondary education. Learners dropping out of mainstream secondary education receive practical training until education is no longer compulsory (PRO, *Praktijkonderwijs)*. However, the majority of 12-year-olds attend the mainstream of secondary education in the Netherlands.

A single target applies to all types and stages of secondary education in the Netherlands, i.e. 'that public-authority education should contribute to the development of pupils with due regard to current ideological and social values within Dutch society and recognising the significance of the diversity of those values' (Eurybase, 2003: 5.3.2.5).

Dutch secondary education encompasses three types: pre-vocational, senior general secondary, and pre-university. Below, we provide some details on the three types.

1. *Pre-vocational secondary education*: ***vmbo*** *(voorbereidend middelbaar beroepsonderwijs)*

   Vmbo lasts for four years. It is a relatively recent type of education first introduced in 1999. Vmbo is a merger of two types of education that had existed before: pre-vocational education (vbo) and junior general secondary education (mavo). Three features are typical of the vmbo curriculum. First there is the inclusion of a table for basic secondary education (Basisonderwijs) common to all types of secondary education. Attaining the targets associated with this table takes learners at least two years. Additional characteristics are the choice for a particular sector (idem) and a particular learning pathway (leerweg) at the end of second year at the earliest. There are four sectors: engineering and technology, care and welfare, business, and agriculture. The four learning pathways are the theoretical programme, combined programme, middle-management vocational programme, and basic vocational programme. After completing vmbo at the age of around 16, learners either enter the job market or move on to secondary vocational education (mbo), which is a form of further vocational education. Learners who have successfully completed the theoretical pathway of vmbo can also opt for promotion to the fourth form of senior general secondary education.

2. *Senior general secondary education*: ***havo*** *(hoger algemeen voortgezet onderwijs) Havo lasts five years and is for learners aged 12-17. Havo is a type of education first introduced in 1968, in the wake of a secondary education act commonly referred to as the mammoth act (Mammoetwet) because of its formidable size and impact. Havo was similar to its British counterpart Secondary modern. Both were geared at lower attainment levels than the levels the traditional grammar school aimed for. Havo consists of a first stage of three years, in which the curriculum of basic secondary*

education is integrated, and a second stage of two years commonly referred to as second phase. The revision of the 1968 secondary education act on 1 August 1998 contained three main implementing regulations that in particular effected senior general secondary (havo) and pre-university (vwo) education. The three decrees were:

- *the Basic Secondary Education (Attainment Targets and Recommended Number of Periods per Subject 1998-2003) Decree which contains details of the attainment targets for basic secondary education and the recommended number of periods per subject;*
- *the Secondary Education (Organisation of Teaching) Decree, which regulates teaching in the different types of school, including the admission requirements and the recommended number of periods per subject;*
- *the VWO-HAVO-MAVO-VBO Leaving Examinations Decree, which regulates the choice of examination subjects and stipulates how examination results are to be determined. (Eurybase, 2003: 5.2)*

The havo curriculum is presented as appendix 1. Havo graduates often move on to institutions for higher professional education (Hogescholen) or Universities of Professional Education, as they are entitled to be called in international contexts by consent of the Dutch Minister of Education, Culture and Science. Havo graduates can also decide to go on to the fifth form of vwo.

3. *Pre-university education*: ***vwo (****voorbereidend wetenschappelijk onderwijs)*

Vwo lasts six years and prepares for entry to one of the thirteen universities in the Netherlands. Learners holding a vwo-certificate can enter higher professional education, as some of them do. There are three types of vwo education: atheneum, gymnasium, and lyceum.
Atheneum is a modern grammar school that generally does not offer Latin and Greek as curricular subjects. As was the case with havo, this type of pre-university secondary education was introduced in 1968 as well.
The oldest type of Dutch pre-university secondary education is called gymnasium. It is a classical grammar school where Latin and Greek are compulsory subjects. Since 1968, the gymnasium learners have the option to graduate in only one of the classical languages.
Lyceum is an integrated grammar school offering Latin and/or Greek as optional subjects.

There are no formal distinctions between the examinations and certificates the learners of the three types receive at the end of their education. Admission to a particular faculty depends on the subjects or *profile* of subjects the learner graduated in. *Vwo* consists of a first stage of three years, in which the curriculum of *basic secondary education* is integrated, and a second stage of three years called *second phase.* The revision of the 1968 secondary education act on 1 August 1998 has effected the *vwo* curriculum in ways similar to how *havo* tables have been effected. Details on the *vwo* curriculum are given below.

In general, secondary schools are comprehensive schools offering all of the three types of secondary education, sometimes in different locations. The increase of large comprehensives has been the result of the government's mergers' policy. Schools were lured into mergers to improve management efficiency and facilitate transfer from one type of education to another within the same school. The mergers' policy had been preceded by a new financing system, in which the schools were granted a lump sum of money to spend as the management thought fit. These two efficiency and deregulation measures preceded the future curricular and didactic reforms to come. The new mode of block-grant financing required the necessary financial and

organisational expertise of school managers. The mergers were often lengthy and painful processes, which involved loss of identity for teachers and learners alike. Despite the financial encouragements of the authorities, some schools decided not to merge. This was for instance the case for quite a few *gymnasium* schools.

In the next section, we will discuss the three types of teacher known in the Netherlands and provide some information on teacher education.

6.4 Teacher qualification in the Netherlands

The respondent teachers in this investigation are all grade-one teachers of English. In order to interpret what this means in terms of tasks and qualifications, we will briefly discuss the three types of teacher qualification in the Netherlands.

There are three teacher qualifications in the Netherlands: *primary, grade-two,* and *grade-one* (or *full-grade*).

Primary teachers are qualified to teach all subjects and groups in primary education. However, the teachers educating the toddlers in groups 1 and 2, and the teachers teaching e.g. English or physical education, usually have specialised in these areas. Primary teachers are trained at Colleges of Higher Professional Education called PABOs (*pedogische academie voor het basisonderwijs)*. PABOs are often part of larger institutions for higher professional education (HBO), often referred to as Universities of Professional Education in international contexts. The minimum entrance requirement is a *havo* secondary school certificate. A typical programme is four years, and includes an increasing amount of teaching practice over the years.

Grade-two teachers are qualified to teach *vmbo* and the first three forms of *havo* and *vwo.* Grade-two teachers graduate in one school subject at Colleges of Higher Education. The Netherlands does not have a dual certification system. Similar to the primary teachers to be, an aspiring grade-two teacher needs to have a *havo* certificate at the least. As was the case with PABOs, the Colleges of Higher Education are often part of larger institutions.

A grade-two course also takes four years. In these years, subject knowledge, skills and understanding are developed, next to didactic expertise. Teaching practice is an essential component, especially during the last two years.

Grade-one teachers are qualified to teach all levels and forms of secondary education.

Student-teachers receive their one-year training at a Graduate School of Education of a university. Entrance qualification is a university master's degree (MA, MSc) in the subject area concerned. In general, student teachers only qualify in one subject. Most of the aspiring teachers have done an orientation didactic course as part of their masters' education. Others already hold a grade-two degree, and have successfully completed their masters.

Part-time teacher education and differentiated programmes have increased since the late 1990s. This is primarily caused by the dramatic shortage of primary and secondary teachers in the Netherlands. The differentiated programmes and the call for more qualified teachers seriously challenge teacher education institutions. In addition, the laws of supply and demand often force schools to temporarily appoint student teachers as grade-one teachers by lack of candidates who have already qualified. Aspiring first-grade teachers increasingly first find jobs, and next register with a college or university.

In the sections that follow, we will first pay attention to English in primary education. Then, we will survey English as a school subject in the first three or four years of secondary education. Finally, we will discuss English as it is taught in the upper forms of *havo* and *vwo* education. We will pay attention in particular to the

curricular and didactic innovations that affected lower and upper secondary education in the Netherlands. The innovations are generally referred to as *Basic Secondary Education* (Basisvorming) and *Second Phase (Tweede Fase).*

6.5 English in primary education

Since 1986, English has become a compulsory subject for the two upper groups in primary education (groups 7 & 8). In this section we will discuss the attainment targets of primary English and the ways in which primary English is monitored.

Attainment targets

In line with the fundamental liberties, the contents of teaching and the teaching methods used have not been prescribed. Nevertheless, attainment targets have been formulated. Schools are expected to organise their teaching in such a way that all of the targets will have been achieved when the learners leave primary school at the age of twelve. The targets define the core curriculum in broad terms and ensure that pupils are prepared for secondary school entrance. More elaborate attainment targets have been formulated for the core subjects of Dutch and Arithmetic and mathematics, by way of intermediate targets and teaching guidelines. At the time of writing, the primary attainment targets were under review.

Monitoring primary school English

CITO periodically carries out standard measurements of proficiency levels in primary education. Periodical measurements of English as a primary school subject took place in 1991 and 1996 (Edelenbos, P, van der Schoot, F., & Verstalen, H., 2000). Both measurements encompassed an inventory of relevant aspects of teaching in year group 8 and a measurement of eight distinctive learner skills. Standards for most of these skills provided a frame of reference that helped to interpret learner achievement. The standards were based on the attainment targets for English as a primary school subject. The most relevant conclusions from the second measurement in 1996 are presented below. The results of the 1996 measurement illustrate how English at primary level was assessed.

In the light of our investigation, we would like to single out two results. First of all, apart from a decrease in the quality of pronunciation, the 1991 and 1996 measurements do not show any serious differences in levels of skills attainment. The measured levels of proficiency suggest consistency. This brings us to a second outcome. The actual levels of proficiency remain disappointingly low, compared with the levels that are targeted for. The standard ‘sufficient’ is only attained by 46% of the learners for listening, 33% for speaking, 25% for speaking vocabulary, 49% for reading, and 34% for reading vocabulary. We will relate these relatively low scores on the four language skills and levels of vocabulary to some of the findings of this study in our discussion chapter.

In the next section we will discuss how English education has been organised in the first three years of secondary education.

6.6 English in basic secondary education

Basic secondary education was first introduced in 1993. It involves a core curriculum in the first stage of secondary education in the various types of school. National attainment targets have been set for the subjects in the core curriculum. The attainment targets are set by law for periods of five years. The targets indicate the expected level of attainment in terms of knowledge, understanding and skills. They

are compulsory minimum standards that schools are expected to achieve by the end of the period of basic secondary education.

During the first two or three years of secondary school the learners are taught a compulsory curriculum of 15 subjects. In the first three years of secondary school, pupils must be taught for at least 1,280 periods of 50 minutes per year. The time spent on educating the core curriculum is expected to take up about 80% of the available teaching time. The remaining 20% of teaching time may be used by schools at their own discretion, e.g. for additional lessons or other educational activities. Schools may, for instance, decide to pay attention to Latin, religious education, mother tongue teaching, pre-vocational subjects or devote extra time to subjects from the core curriculum, private study or vocational orientation. The schools are free to differentiate in time and extent for each learner or year group.

Tables have been published that indicate the recommended number of 50-minute periods to be spent on each subject. There are three new subjects: *technology*, *IT studies* and *life skills*. The number of periods shown for the arts (visual arts education, music, dance and drama) and physical education is the prescribed minimum number of periods. A third modern language is compulsory in the first three years of *havo* and *vwo*, and the *gymnasium* curriculum logically includes Latin and Greek.

Attainment targets

With the introduction of basic secondary education in 1993, a core curriculum comprising attainment targets became compulsory. The Secondary Education Act specifies that the targets are to be redefined every five years. The 1998-2003 curriculum (The Dutch Ministry of Education, Culture and Science/ Ministerie van Onderwijs, Cultuur en Wetenschappen, 1998b) encompassed:

> *general attainment targets, referring to eight cross-disciplinary themes (1) and five learning skills (2-6).*
>
> *specific attainment targets for the fifteen compulsory subjects. For the foreign languages English, French, and German there are twenty-three targets, divided over the five fields Reading skills, aural and observational skills, conversational skills, writing skills, and language and culture.*

General attainment targets

The cross-disciplinary themes have been derived from societal phenomena, such as the relation between man and nature, the significance of technological development for society, art, culture and emancipation. They are diverse, as the following three examples may show: “recognizing and dealing with one’s own standards and values and those of other people” (1.1), “acting with due regard for personal and general safety within one’s own environment and in traffic” (1.5), and “the achievements and possibility of art and culture, including the media (1.8) (Ministerie, 1998b, 1.1)

The five learning skills that are highlighted are:

> *Learning to do: geared at the further development of certain scholastic skills, such as “comprehending written and spoken Dutch and English”(2.1) or “using computers (2.7).*
>
> *Learning to learn: strategies to improve the learning process, such as “assessing information on the grounds of reliability, representativeness, and usefulness, as well as processing and utilizing information” (3.1), “making choices on a considered basis”(3.4), and “developing a personal opinion on the basis of argumentation” (3.7).*

Learning to communicate: directed at further development of social and communicative skills, such as "respecting elementary social conventions (4.1), "conversing and working as part of a team"(4.2), and "presenting oneself and one's work"(4.7).

Learning to reflect upon the learning process: having learners learn to be analytical about and to control the learning process, by reflecting on their own performance. The three topics mentioned are: "planning work" (5.1), "monitoring the learning process"(5.2), and "making a simple product and process evaluation and drawing conclusions for future application"(5.3)

Learning to reflect upon the future: having learners learn to be analytical about options for future prospects and interests by reflection on their own performance. Among the five skills mentioned are: "listing personal capabilities and interests" (6.1) , "the role and significance of the knowledge, insight and skills acquired at school" (6.3), and "ways of using one's leisure time"
(6.5).

The Dutch ministery of Education, Culture and Science acknowledges that a gradual swing is needed to integrate general attainment targets into the subject-specific attainment targets and everyday teaching. The development of basic secondary education is considered to be in line with the curricular and didactic reform that is to take place in upper secondary education. A gradual innovative swing is expected to be achieved if teachers, primary and secondary alike, focus on:

- *skill-focused education that aims for a balanced development in the student of knowledge, understanding and skills;*
- *guided education that focuses on a more active role for the student, from teacher-directed towards student-directed education;*
- *differentiated education that recognizes the differences between students and gives these differences their due;*
- *coherent education in which teachers individually and together discover and define what contributions their subjects can make to the collective responsibility they carry for the development of their students.*

(Ministerie, 1998b:8)

Attainment targets for basic secondary English

This section is based on the targets set for 1998-2003. The first version of the attainment targets differs from the second version in a number of respects. As compared to the 1998-2003 targets, the 1993-1998 version had:

- a less elaborate general objective, geared to classifying communicative situations;
- no explicit references to the ways in which the subject-specific targets are related to the general attainment targets;
- 19 targets, instead of 23;
- a division into four domains:
  a) *communication skills (15 targets)*, subdivided into the traditional subdomains reading (5), listening (3), speaking (5) and writing (2).
  b) *compensatory strategies and techniques (1 target)*
  c) *socio-cultural competence (1 target)*
  d) *orientation to learning foreign languages (2 targets)*

Although the 1998-2003 targets are more elaborate and specific on communicative language proficiency, the first version of attainment targets is still traceable in its revised counterpart.

The attainment targets published and monitored by the Ministery of Education, Culture and Science do not specify any desired levels of achievement "… each school is free to determine the achievement level for its students. The assumption is that a teacher will 'place the bar as high as possible, depending on the potential and interests of the students.' (Ministerie, 1998b:8).

The targets for English are similar to the targets for French and German. The attainment targets for modern foreign languages are preceded by a comprehensive general objective and by a section that lists which of the general attainments targets should at least be achieved in the foreign language curriculum. The general aim for foreign languages is presented below, followed by examples of more specific attainment targets for the foreign languages.

> **General objective**
>
> *The object of education in the German, English and French languages is to enable pupils:*
> - *to use the languages in question in to communicate as effectively as possible in various situations which commonly occur and/or are relevant to the use of language;*
> - *to appreciate the scope afforded by the languages in question. by developing a feel for effective communication and by taking pleasure from communicative situations;*
> - *to employ strategies in various situations for making the best possible spoken and written use of their limited knowledge of the languages in to use the languages in question to communicate as effectively as possible question;*
> - *to acquire a degree of insight into the societies and the cultures of the main countries to which the languages in question are native;*
> - *to acquire a degree of insight into the role and importance of the languages in question in relation to social interaction;*
> - *to acquire knowledge of and skill in the use of the languages in question with a view to facilitating decisions regarding further education, subsequent employment and social activity, including recreational activity.*
>
> (Ministerie, 1998b:21)

The foreign language targets for basic secondary education are divided into five fields: *reading skills (A: six targets); aural and observational skills (B: 5 targets); conversational skills (C: 5 targets); writing skills (D: 3 targets),* and *Language and Culture (E: 4 targets).*

Monitoring basic secondary English

The Dutch Education Inspectorate carried out a comprehensive study in which the first five years of basic secondary education (1993-1998) were evaluated. The results of the 1998 investigation of the Education Inspectorate (Inspectie van het Onderwijs, 1999b) may help to interpret the English knowledge, skills and understanding of the fourth-form learners of the present investigation.

The investigation was based on a sample of 119 secondary schools, and the observation of 600 lessons. In addition, interviews were held with individual teachers, departments of English, and school managers.

With a touch of political understatement, the Education Inspectorate concluded that five years after the introduction of English in basic secondary education there was room for improvement. Too few schools managed to realise all of the attainment targets. English teachers and departments of English were generally not preoccupied

with developing a programme in line with the features and targets of basic secondary education.

The school inspectors have also monitored the teachers' general didactic approaches. The Inspectorate found that the quality of the areas effective class management and the ability to differentiate often did not reach the expected norm. A positive exception is the adequate way in which most of the teachers manage to challenge and stimulate the learners, and offer them structure and a safe environment. 77% of English teachers achieve the norm set for the indicator *pedagogical approaches* (*pedagogisch handelen*). Similarly, the learner results on the national basic secondary tests for English are good, with 99% achieving the minimum requirements. The results for French and German do not differ substantially from the outcome for English.

The results led to the following recommendations:

6.1 Teachers

- *It is to be recommended that teachers use the course materials as guidelines and not as prescriptions. Teacher should deliberate plan their education based on the attainment targets. In initial and in-service teacher training, attention should be paid to the variety and choice of course materials in relation to the attainment of the targets for basic secondary education.*
- *English programmes in basic secondary education require didactic impulses: less ex-cathedra teaching and a greater variety in didactic procedures that are primarily skill-based. In particular, more attention should be paid to the acquisition of oral proficiency and computer use.*
- *An essential part of communicative language teaching is the use of English as the medium of communication. It is to be recommended that teachers of English make arrangements on target language use and realise them in their day-to-day practices.*
- *It is to be recommended that teachers of English agree on what is to be to dealt with and what is to be skipped. Likewise, arrangements related to the methodology of teaching and learning English are of importance. More cooperation and consultation with colleagues of French and German on these matters is required.*

6.2 School management

- *Secondary education does not or hardly ever make use of the knowledge of English the learners have acquired in primary education. This knowledge is considered to be too diverse in quality. It should be considered whether increased differentiation can help to keep the English programme challenging for all of the learners. It would even be better to make arrangements with primary education and become more knowledgeable of how English is dealt with in primary education.*
- *School managers in junior secondaryvocational education ((i)vbo) should ensure that the hours pent on English come closer to hours recommended in the tables. School management can decide to raise the number of periods taught in a year or decide to finish attainment of the targets at a later stage.*
- *School managers should strive to have English taught in separate subject classrooms. The appearance and facilities of other classrooms are often inadequate. Better and more direct access to computer facilities is strongly recommended.*
- *School managers should stimulate and enable in-service teacher training. It particularly concerns training directed at varieties of*

didactic procedures, skills, particularly oral skills, and the use of computers in support of foreign language acquisition.

6.3 Government

- *It is not realistic to expect special schools for junior vocational secondary education (ivbo's) to adequately deal with the attainment targets. An example of a target that could be skipped is "extensive reading of longer texts".*

(Inspectie van het Onderwijs, 1999b: 46-7)

There definitely is room for improvement for English as a school subject in basic secondary education. Teachers are recommended to pay more attention to the attainment of the predominantly communicative targets of junior secondary education, be less dependent on the course materials, (have the learners) speak more English, and extend their didactic repertoire in cooperation with others. School managers are expected to stimulate, and perhaps even more importantly, facilitate their teachers to bring about the recommended changes. The young learners, with sometimes hampered communicative competence, are in at the centre of the recommendations of the Dutch School Inspectorate. Yet, the third-form havo and vwo learners are to face even bigger challenges when they move on to the fourth forms of their types of education and enter the next stage of secondary education.

6.7 English in the *Second Phase*

Senior secondary education encompasses the two types that prepare for higher or academic, that is *havo* (forms 4 and 5) and *vwo* (forms 4, 5 and 6). The period follows on from basic secondary education and the curricula of the third forms. Education in these upper forms of secondary education is generally referred to as *second phase.*

The renewed second phase was the result of a combined curricular and didactic innovation meant to be introduced in August 1998. By that time, only a limited number of schools was ready to implement the organizational and didactic reforms it aimed to bring about. That is why the Minister of Education, Culture and Science decided that the official introduction would be postponed till August 1999. 123 schools implemented the proposed changes in 1998.

Two types of changes are typical of the innovations of upper secondary education in the Netherlands: curricular reform and didactic reform.

Curricular reform

Up to the introduction of the *second phase*, the schools and learners had been relatively free in choosing combinations of six (havo) or seven (vwo) school subjects for graduation. In the *second phase*, the learners must opt for one of the four so-called curricular *profiles* (*profielen*):

- *science and technology,*
- *science and health*
- *economics and society*
- *culture and society*

All of the profiles consist of a *common* component of subjects, a *specialised* component and an *optional* component. Dutch and English are always part of the common component. The ministry has produced tables that set the yearly study load for *havo* and *vwo* learners at 1,600 hours (equivalent with 40 weeks of 40 hours), at least 1,000 hours of which should consist of teaching during school time. The schools

are free to divide this load over the two or three years of *second phase* education. The standard study loads of *havo* and *vwo* are:

HAVO

- 1,480 hours for the common component;
- 1,160 hours for the specialised component;
- 560 hours for the optional component.

VWO

- 1,960 hours for the common component;
- 1,840 hours for the specialised component;
- 1,000 hours for the optional component.

The curricula of havo and vwo are presented as appendices I and II.

Didactic reform

Second-phase reform also concerned the ways in which the learners were taught and expected to learn. The educational reforms introduced in the upper forms of havo and vwo involved a new approach to teaching, with the emphasis on knowledge acquisition and active learning through independent study rather than knowledge transfer (teaching). Therefore, the teacher's role was meant to shift from that of instructor to supervisor and coach.

The metaphor of *Studiehuis* (House of learning) was introduced to refer to the didactic reform, which affected the schools physically, by way of facilities for learners to work and/or learn on their own, as well as didactically, by means of the experimentation with methodologies that aimed at increasing learner autonomy by stimulating active and self-regulated learning (*actief en zelfstandig leren*). It involved a shift from traditional *teacher instruction-learner reproduction* patterns to modes that asked for more learner involvement, activation and self-regulation. Exactly how this was to be achieved has been open to debate ever since the introduction of the renewed second phase. In our discussion chapter we will relate some of the results of the monitoring projects of the second phase to the findings of our study.

Aims of second phase reform

Van den Akker & Bergen (1997) discuss the arguments used to justify the need for the massive curricular and didactic reform of in upper secondary education. The authors translate these justifications into eight aims. The second-phase innovations are meant to:

1. *logically link up with the innovations of basic secondary education;*
2. *guarantee a better preparation for tertiary education, as strategic study skills are increasingly paid attention to;*
3. *prepare for social changes, especially the tendency towards life-long learning in a knowledge-intensive society and flexible labour market;*
4. *acknowledge the consumer role of adolescent learners in today's schools;*
5. *tackle the motivation problems of pupils as observed by teachers;*
6. *apply recent constructivist theoretical views of the dynamic relationship between learning and teaching;*
7. *take into account the rapid development of knowledge within the scholarly disciplines that underlie the school subjects by reformulation of targets and contents, in which learning skills rather than the extension of fast-dating knowledge is emphasised;*

8. *take account of the tempestuous developments in the field of information and communication technology (ICT), which requires different emphases in learning targets (especially more concentration on skills) and also offers a promising and stimulating learning environment for the realisation of learning targets.*

(Van den Akker & Bergen, 1997:120/1, translation mine)

Van den Akker & Bergen argue that some of the arguments in favour of the *Studiehuis* concept have been heard for time immemorial, particularly those of learner consumerism and learner motivation. The field of education did not initially object to the rationale of the need for change. Only few teachers and school managers were negative from the start. It was the result of an intensive campaign by PMVO (process management of secondary education), which informed teachers about the proposed changes, stimulated discussion and reflection, and challenged schools and teachers into piloting some of their first ideas. The Graduate School of the University of Nijmegen was among the institutions active in this field by setting up two successful school networks related to integrated study-skills education (*geïntegreerd studievaardigheidsonderwijs*) and active and self-regulated learning (*sturen en activeren*). Both projects successfully stimulated peer coaching, reflection and professional development.

Because curricular and didactic innovations were to take place simultaneously and educational staff had to implement the reform often without being adequately facilitated, it was envisaged that implementation would be no easy matter. Van den Akker & Bergen (1997) expressed their concerns two years before the actual implementation was to take place nation-wide.

> *There is obvious support for these fundamental innovations. However, on the basis of empirical data from comparable fundamental innovations in education, it can be predicted that there will be big operational and implementational problems in relation to the learning-to-learn process of the pupils, the testing of the targeted skills, and the acquisition of the necessary methodological repertoire by teachers. We must realise that there is a wide gap between the ideals of the Study House and the 'average' teaching methodology currently in use. …. We are not being unduly pessimistic, but rather realistic if we take into account that, in general, the fate of attempts at educational change is rather sad, if, at least, the actual effect in concrete educational terms is anything to go by. Many attempts have stranded on the stubborn rocks of habitual practice.*
> (Van den Akker & Bergen, 1997:121)

We will inevitably return to Van den Akker & Bergen's alleged prophesies in our final chapter, when we interpret the findings of our qualitative study in the context of second-phase reform. Now that we have discussed the intended curricular and didactic changes in the upper forms of secondary education, we will look more closely at second-phase English.

Attainment targets for second phase English

The examination programmes and attainment targets for second phase subjects were published in 1998 (Ministerie van Onderwijs, Cultuur en Wetenschappen, 1998a). However, measures were taken by the Education Secretary to alleviate pressures in educational practice from as early as the school year 1999-2000.

The attainment targets are similar in contents and lay-out to the first version of the attainment targets for basic secondary education (1993-1998). The second phase targets are also characterised by a division into domains. Fifty (vwo) or forty-three (havo) attainment targets have been divided over the six domains of *Reading skills*,

Listening skills, *Speaking skills*, *Writing skills*, *Literature* and *Orientation on further studies and careers*.

The *havo* and *vwo* examination programmes for "English language and literature" are textually alike. Yet, there are differences between the two programmes. Whereas the language skills of vwo-students of English are assessed according to specification level 4, havo-students of English are assessed on the basis of specification level 3. In addition, havo-learners do not have to attain targets 25 & 26 (speaking skills: instruct and give and justify an opinion), 39 (general skills: writing a report in English), and 44-47 (the subdomains literary notions and literary history).

The programmes encompass school examinations over a period of two (*havo)* or three *(vwo)* years, and a nation-wide final examination of reading comprehension. The certificate grade for English is made up of the final examination score, averaged by the common mark of all of the school examinations sat by the school leaver. Examination marks are given in one decimal, whereas the certificate mark is a full mark. The 1-10 scale is common to secondary education in the Netherlands, with 10 being the maximum score. A score of 5.5 is usually the caesura of sufficient and insufficient achievement. More details on the havo and vwo curricula can be found in appendices I and II.

As an illustration of the second phase attainment targets, we will present and discuss the domain of writing skills.

> ***Domain D: Writing skills***
> ***Subdomain: Language skills***
>
> *The candidate is able to:*
> *29 thank, apologise, congratulate, and invite.*
> *30 give and ask for information.*
> *31 give an opinion and ask for one.*
> *32 describe someone or something.*
> *33 express and ask for feelings, interests, and preferences.*
> *34 comment and pass judgement.*
> *35 plead and complain.*
> *36 to express and justify a point of view*
>
> *Attainment targets 29-36 are assessed on the basis of level 4 specifications of writing skills for vwo learners and level 3 specifications for havo learners.*
> (Ministerie, 1998a, translation mine)

The targets for writing skills invite two comments. It is first of all remarkable that the attainment targets for the second phase are less detailed than the ones for basic secondary education. Secondly, we would like to argue that the formulation of the eight targets for second-phase writing skills, which are similar to the goals formulated for speaking skills, are neither helpful nor practical for users due to their lack of specification.

Imagine the following e-mail exchange between two adolescents in English: *I hate school, because it sucks. What about you?* The grammatically correct and appropriate response of the adolescent's willing interlocutor could then be: *I think it sucks, too.* Without any detailed level specifications, the two of them may have attained targets 31, 33 and even 34, if one is generous. Another written exchange might include the following section: *In your letter you asked me to explain why I preferred Schopenhauer to Nietzsche. Well, I'm more attracted to Schopenhauer's philosophy, because unlike Nietzsche he…* In the sentences that follow, this learner might well illustrate that (s)he is able to give and ask for information or an opinion (30,31) as well as able to describe and comment on Schopenhauer and/or Nietzsche and their respective philosophies (32,33,34, and 36). Therefore, let us turn to the level specifications for more clarity in these matters.

Five attainment levels have been indicated for foreign languages in the second phase, of which only level 3 and 4 are used (Minsterie van Onderwijs, Cultuur en Wetenschappen, 1998: 253-260). Unfortunately, the level indications do not seem to be specific enough to be practical in determining whether and when a learner has attained particular targets. In general, level 4 asks for more knowledge and more skills than level 3. However, exactly what this knowledge and these skills constitute has not been specified. This is basically left to the interpretations of the users, most notably the teacher and the learners.

Targets 29-36 were part of the subdomain language skills. There are two more subdomains of writing skills: practical situations and general skills.

> *Subdomain: Practical situations*
>
> *37 The candidate has participated in correspondence projects, if needed by way of telecommunication.*
>
> *Subdomain: General skills*
>
> *38 The candidate is able to use the facilities offered by word processing when writing.*
> *39 The candidate has written a written a report in the target language, for the benefit of a project or another school subject.*

First of all, havo-learners do not have to attain target 39. Even though targets 37-39 lack detail, they seem directional in stimulating correspondence projects, computer use, and the use of English in reports, even if it only concerns one report. Other guidelines are presented under the heading of assessment. Teachers are expected to assess and/or evaluate learners' writing skills in terms of content, structure and appropriateness .

Assessment

The following aspects need to be considered: content (completeness, originality, clarity), structure, and correct and appropriate language use.

The message seems clear here. The use of correct grammatical structures is not the only aspect that needs to be considered when assessing writing skills. Yet, exactly how the components are weighted is left to the discretion of the teacher.

Our data will show how the three teacher respondents will go about assessing writing ability and what consequences this may have for their learners.

A formidable task

Grade-one teachers face formidable tasks. They have to integrate the second phase reform with its broadly formulated attainment targets with their everyday practice as teachers of English. The didactic changes involve more attention to communicative competence and to fostering autonomy in the learners. Exactly how these changes are to be implemented is only indicated in broad terms. Moreover, teachers largely depend on the course materials and their school management for the necessary guidance and facilitation.

In the years that preceded the introduction of the renewed second phase, we observed that the innovations were introduced top-down, with little consultation of practising teachers. We regretted the absence of classroom and teacher data and knowledge about important constructs that are relevant to the subjects the teachers teach. We felt the following six aspects were generally missing at the start of the renewed second phase:

- *havo* and *vwo* classroom data based on planned research designs;
- exploration of how the innovations relate to teachers' core beliefs;
- exploration of how teachers define core constructs of the subjects they teach and how they subsequently put these into practice;
- evaluation of the role of assessment and evaluation in second phase education subject-specific specifications of good practice in second phase education;
- fundamental discussion of the relationship between domain-specific core constructs and the second phase innovations, with the question in how far the innovations are conducive to fostering autonomy as well as communicative ability in the adolescent learners.

We will end this chapter with a brief summary and then move on to our first, predominantly narrative, data chapter. That chapter is based on the interviews we had with the three respondent teachers just before the start of the school year

6.8 Summary

In this chapter we concentrated on the educational context of the present investigation.

We first mentioned the two legislative parameters of education in the Netherlands, that is *freedom of education* and *compulsory education*. Then we discussed how Dutch education has been organised, with attention in particular to the three types of secondary education: pre-vocational secondary education (vmbo), senior secondary education (havo), and pre-university education (vwo). This was followed by a discussion of the three teacher qualifications in the Netherlands: primary, grade-two and grade-one. Then we concentrated on the subject of English, which we discussed in relation to primary, basic secondary and second phase education.

Large-scale monitoring of primary and basic secondary English had shown that it proved to be difficult to attain the predominantly communicative targets that were set. From 1998/1999 onwards, senior secondary education had to implement a massive curricular and didactic reform, which is generally referred to as *second phase.* The didactic focus for havo and vwo learners in the fourth forms was on learning to learn and on fostering learner autonomy. On top of that, the learners and teachers also had to face a large-scale curricular reform. Amidst the hustle and bustle, our respondent teachers attempted to teach their learners how to communicate in English. We suggested that the attainment targets that were set for second phase education were neither helpful nor practical for the teachers and learners. Guidance and support was generally offered by a host of course materials that had not been used before or by progressive school managements having clear ideas on how to successfully implement the reform of the second phase.

The four chapters that follow will report on our teacher data.

CHAPTER 7: THREE STORIES TO TELL

7.1 Introduction

In the first of our four data chapters, we would like to introduce our respondent teachers. The title suggests that stories will be told. By combining factual information with relevant quotes from the interviews, we will attempt to have the three teachers directly convey their core beliefs and concerns related to learner autonomy, communicative language education and language assessment and evaluation. One learns about their histories, beliefs and concerns related to teaching and testing. Most of the information will be narrated by the teachers themselves, by presenting quotes of relevant sections from the first interviews we had at the beginning of the school year. The quotes were selected on the basis of what the teachers regularly mentioned as essential regarding the three main constructs of the investigation throughout the year of data collection.

The option for this particular mode of narration is an attempt to solve a dilemma that faces most qualitative researchers. This dilemma is how to report on the often extensive qualitative data in effective and convincing ways. The direct quotes in this chapter should help to ensure that the stories that are told in this chapter are not just tales that sprang from the mind of a single researcher, who has, of course, histories, beliefs and concerns of his own. Readers should ideally be allowed to use their own frames of reference in interpreting what the three teachers believe in and what concerns them most in teaching and testing.

In this chapter we will first introduce the three respondent teachers. We will go into the reasons why the three teachers were asked to participate in the investigation. In addition, we will provide some details on their careers and personality characteristics as they became known in the course of the year of data collection. Finally, the respondent teachers' schools will be discussed. After this general introduction, the three teachers will have the floor. We will successively listen to *Joy, the budding professional, Mark, the literary master* and *Pete, the project man.* The teachers address the issues mentioned in the first interview guide, which can be found in the appendices.

The issues and question categories of the interview guide reflect the objectives and research questions of the present study. Thus, Joy, Mark and Pete will confide to us some of their *formative experiences and core beliefs concerning teaching and testing*. Next, the three teachers will explain how they interpret the three main constructs of the present investigation, that is *learner autonomy, communicative language education*, and what they consider to be effective *assessment and evaluation* of the learning results. The section on assessment and evaluation has been subdivided into the *knowledge, skills and understandings* the respondents consider necessary for their learners to master, their interpretations of an *effective written English language test*, and their perceptions of *washback*. The one but last section is about the plans and intentions the respondents have for the school year that is about to start. Finally, we will briefly refer to the language tests that Joy, Mark and Pete were to select for discussion in the course of the school year. The data on these tests will be presented in the three chapters to come.

We will start with some information on Joy, Mark, and Pete.

7.2 Introducing Joy, Mark, and Pete

In 2.6.1 we discussed the criteria used to select three respondent teachers for this study. The three teachers that best matched these criteria, readily accepted our request to join them for a year in one of their fourth forms. They were expected to be

precise and accurate classroom managers, used to developing and constructing their own course materials and tests. In their didactic approaches they were supposed to aim for high levels of knowledge of and skills in English geared at communication. They appeared to be appreciated as teachers by their school managers and learners. They had also shown to be genuinely interested in the individual learner. All of the three teachers were known to be explicit about their core beliefs in teaching and testing.

One of the three teachers of English is a former teacher trainee of ILS. We have called her *Joy*, because she had given us joy in the course of her teacher education. Then there is *Mark*. He is an experienced teacher who has been a school supervisor of teacher trainees for many years. Mark is critical of the more recent curricular and didactic innovations. I have called him Mark, because, just like the great apostle, he has a gospel of his own. And finally there is Pete. He is also an experienced teacher of English. Unlike Mark, though, Pete feels that the second phase didactic innovations seem to fit in well with the didactics he has developed since the start of his career. He is called Pete, because he reminded us of a good-natured, but slightly rebellious character out of Dutch youth reading called “Pietje Bel”.

We would like to stress that despite differences between the three teachers, they have all come up to our expectations as good-practice teachers as we had initially defined the term. Most of the rhetoric of the first interviews was validated by the data in the course of the year. The levels of English of the language tests generally did not disappoint us, and could sometimes be called academic. Perhaps even more importantly, the three teachers did not only prove to be competent teachers, but also appeared to be educators genuinely interested not only in their learners’ academic achievements, but also in their moral development and well-being.

In three consecutive sections, we will present additional background information on the three respondent teachers and the schools at which they teach. The respondents’ ages are their ages at the beginning of the year of data collection.

7.2.1 Joy

Joy is a 31-year-old teacher of English who has been teaching since 1992. She started her career in junior secondary education immediately after finishing a four-year degree-two course at one of the colleges for teacher education the Netherlands. Unlike most of her former fellow students at the college, Joy was a university graduate when she started doing her degree-two course. After a successful first year in English language and Literature at a Dutch university, she had pursued her studies in English and Italian translation and got her MA degree.

In 1993 she was employed by her present school. Given her background and competences, she was soon teaching upper secondary classes. School management considered her competent to teach in upper secondary education, even though she was not formally qualified. She applied for the one-year teacher training course at our graduate school of education to qualify as a degree-one teacher. To this end she took up a half-time sabbatical at her school. Despite her half-time job as a teacher, Joy managed to finish the full-time degree-one course in six months. Joy had qualified as a degree-one teacher just before the start of the school year in which the data was collected.

Joy is not only a diligent and intelligent worker, but also a teacher willing to extend her professionality in areas relevant to the teaching and learning of English, from class management to assessment and evaluation. After a school visit in the course of her degree-one teacher education, she was considered a serious option to be one of the three respondent teachers in our investigation. She appeared on all of the lists of good-practice teachers we had constructed. Despite her age and relatively limited experience, she widely met our selection criteria for good-practice teachers.

Joy is a learner-oriented teacher who is critical and reflective. She is precise and detailed in whatever she undertakes. Joy is not easily discouraged by paper work

and had a prominent role in constructing the study guides and programmes for summative assessment and graduation (PTA) for the 4th formers at her school. Joy is a personality presenting her beliefs in a balanced way. She tends to be a little less outspoken than Mark and Pete in presenting her views, beliefs and convictions. Pete is a colleague of hers at the school she is working at. Joy and Pete have cooperated closely since 1993.

Joy's school is a large comprehensive, offering all of the types of secondary education. In the 1990s the Dutch government had stimulated secondary schools to merge with others so that the full array of secondary education could be offered in one institution. That is why Joy's school has five different locations. The location where she was employed, primarily housed the upper secondary forms of *havo, atheneum*, and *gymnasium*. In the year of data collection, Joy had no junior secondary classes.

In the next paragraphs, we will introduce Mark a little more elaborately. He differs from Joy in more than one way.

7.2.2 Mark

Mark is a 47-year-old teacher who has taught English at the same grammar school for 23 years. He started teaching there immediately after his graduation as a Master of Arts in English and American Language and Literature. The academic title befits Mark. He truly is a master who has specialised in English and American literature. Mark often expresses his appreciation and love of poetry and prose as well as of the arts in general. He regrets that over the years his role as an educator and a master of the teaching and learning process has been reduced to being an organiser, technical instructor and assessor. Of the three teachers, Mark is most critical, and at times prophetic, about the recent curricular and didactic innovations of basic secondary education (*Basisvorming)* and second phase *(Tweede Fase*).

In his first years he predominantly taught junior secondary classes. His first appointment concerned three first forms of 21 pupils each. Gradually, he started teaching upper secondary classes. Initially, he only had one colleague in the English department. Both of them valued their autonomy as teachers. They agreed on essentials and accepted one another's different approaches. He frequently refers to these times as 'the good old days'. Now, the English department has grown with the size of his popular grammar school, and consists of five degree-one teachers of English. In the year of data collection he cooperated with a female colleague of his on matters concerning the 4th-form syllabus.

Mark goes for excellence and perfection, and generally appreciates people striving for the same. He is proud of his grammar school and his talented learners. After some four years of teaching, he became a school supervisor of student teachers doing their degree-one teacher education. Since 1996 we have regularly met in the researcher's capacity as a teacher trainer. As a school supervisor Mark has proved to be particularly good with teacher trainees who aim for excellence and who have gradually come to share some of his core beliefs. More than once, Mark has proved to be a true professional, both as a teacher and as a school supervisor of teacher trainees. Mark is an intelligent and close reader and his often strong views are always well-phrased. In the first interview we had, Mark easily fell into dramatic monologue concerning matters that were close to his heart, with metaphors and all. After a close and deliberate reading of the interview guide, Mark's bottle popped spontaneously and out poured his core values and beliefs.

Mark's grammar school has grown and grown over the years. His school had never merged with schools offering the other types of secondary education. Mark's grammar school has recently been housed in a renovated larger location.

Finally, there is Pete. He contrasts with Mark, not so much in experience as in his views, core beliefs and personality characteristics.

7.2.3 Pete

Pete is 46. He has been a teacher of English at Joy's school since 1978. He started there as a part-time teacher in junior secondary education with six periods a week. Pete had then just finished his *kandidaats*, a former Dutch university qualification comparable to a Bachelor's degree. From that moment onwards he combined his academic studies of English Language and Literature with teaching. His part-time job was gradually extended to 32 lessons per week.

In his early teaching years, Pete claims to have been educated by two progressive colleagues. He gradually came to share their beliefs on teaching, learning and testing. Pete did his MA in 1981, having specialised in English linguistics. In that year he accepted an additional teaching post at another school, where he stayed until 1983. The headmaster of this school had invited Pete to come and work there, allegedly because he wanted Pete to bring about progressive change in his traditionally-oriented English department. After this part-time outing, Pete resumed his full-time job at his present school.

Pete appreciates his school partly because of the teacher autonomy given to him by the school management and the generally pleasant and agreeable learners. Pete is an active member of commissions, organises excursions and foreign trips and is one of the driving forces of the yearly school leavers' gala dinner. Pete is a great improviser, who is not easily discouraged by some extra work. Besides, he seems to cope well with the usual demands and stress of the teaching profession.

We had not met Pete until the spring of 1999, when some of Joy's lessons were observed and discussed. From that moment onwards, he became a serious candidate on the lists of preferred good-practice teachers. We came to appreciate his role in supporting and stimulating Joy when she started teaching upper secondary education classes. His teaching seemed sound and stimulating. We were also struck by his attitude towards the curricular and didactic innovations he was to implement in the school year to come. Pete seemed to lack the usual wariness and cautiousness towards the reform that was typical of many of the good-practice teachers we had pre-selected so far. It would also be a challenge for us to study the role of collegial cooperation in language testing in more detail. After checking with Joy in how far she felt Pete met the good-practice criteria, we decided to ask Pete whether he was willing to participate as a respondent in the investigation. Pete's participation was also justified by a pragmatic reason. Because the time for data collection would be limited and the schools required travelling, it would save time and energy if two teachers were employed at the same school.

Pete is a colleague of Joy's at her large comprehensive. They both teach upper secondary *havo*, *atheneum* and *gymnasium* classes at the same location. Pete and Joy have been cooperating ever more closely since 1993.

This is where our general introduction of the three respondent teachers ends. In the remaining three sections of this chapter, Joy, Mark and Pete hold the floor. They will first tell us about *formative experiences* in teaching and testing since the start of their careers. In describing these experiences, the teachers invariably express their *core beliefs* related to teaching and testing. After that, the teachers will present how they define and interpret conceptions of *learner autonomy* in formal education. The section that follows is on the respondents' interpretations of *communicative language education.* The next part is on the third construct of the present investigation, i.e. *language assessment and evaluation.* Data on this final construct will be presented in three parts. First the teachers describe what *essential knowledge, skills and understandings* they wish their learners to master in the course of their secondary school careers. Next, they will mention what they consider to be the characteristics of an *effective English language test*. Finally, they will present their *interpretations of test washback* on the teaching and learning process.

7.3 Joy, the budding professional

In this section, we will present how Joy responded to the questions and issues raised in the first interview we had at the beginning of the school year. The initial question for all of the three respondents was to describe their experiences with language testing from the start of their teaching careers up to the moment of speaking. The question triggered important formative experiences in teaching and testing, which after analysis appeared to be inextricably bound with their core beliefs. This brings us to the first section of our report, that is Joy's *formative experiences in teaching and testing and core beliefs*.

7.3.1 Formative experiences and core beliefs

Joy started her career as a junior secondary teacher at a secondary school in the mid-east of the Netherlands in 1992. In 1993 she was appointed at her present school in the east of the Netherlands. Here she became a colleague of Pete's. In the school year 1996/7, she started teaching the upper forms of secondary education. The formative experiences she mentions at the beginning of the first interview are related to these two schools. Yet, these experiences were not the only ones mentioned by Joy. When she was a university student of English and Italian translation, Joy offered a cousin of hers to give some extra lessons of English. The question to define the term *insight*, caused her to mention a key experience that appeared to formative in her teaching career.

> *Yes, I know of myself that insight didn't seem to hit me until after secondary school. Not until I decided to take on a language myself. At secondary school I was used to studying course books and preparing for tests in the ways indicated in the classes. And then I decided to study English, just because I really liked the literature, even though my grammar was just so, so, and my fluency even worse.*
> *For listening I had a C average. Once I started studying at the university, I bravely told a cousin of mine "Sure, I'll tutor you." Only then did insight really come when I had to explain matters, and I also had to examine my own way of looking at things.*
>
> *I: Yes, I can relate to that. In the course of your studies you apparently discover the way a language really works.*
>
> *J: Yes, exactly.*
>
> *I: And when that process gets going it actually never stops.*
>
> *J: No, it doesn't and in fact that's something we haven't yet paid much attention to in our classrooms. ..... Yes, so that actually presents us with a real challenge to try and bring it about, don't you think?*
>
> *(1-1-*27)

The moment Joy was in a position to explain to a learner how particular aspects of English worked, she started developing insight in the learning and teaching of English. After her MA in English and Italian translation, she took a four-year course to be educated as a degree-two teacher of English. Afterwards she got her first job as a qualified teacher. She started teaching junior secondary classes and remembers that her colleagues strictly adhered to the course materials. No sidesteps were allowed. When she changed schools in 1993, most of Joy's new junior secondary colleagues appeared to hold similar views to the colleagues she had met at her first school. Yet, there were a few colleagues who were used to making course materials on their own.

J: Yes, at my first school they were really focused on the textbook, on Unicom. And you really weren't allowed to deviate from that book in any way. Whereas at my second school, although I honestly have to admit that there was a lot of discord, some teachers also strictly adhere to the textbook. But others try to use their own materials along with the book.

I: OK. And that mainly has to do with individual differences?

J: Yes, yes. And also with the way in which you approach the subject, especially testing.

I: Right, yes. You actually also state an interesting difference between junior secondary teachers and upper secondary teachers. Was that the case in your first school as well?

J: No, they were more in line with respect to that issue, as far as I can remember.
(1-1-12)

It is important for Joy that colleagues cooperate with one another. She very much supports exchanges of ideas, information, lesson materials and tests. Joy firmly believes that teacher cooperation helps to establish logical connections and links between primary and secondary education (1-1-5) and between junior and upper secondary education (1-1-9;1-1-10). One of the essential arrangements colleagues have to agree on is *what* to test at *what* particular time. When she started working at her second school, she was disheartened that most of her colleagues did not seem keen on teacher cooperation either.

J*: And the junior secondary teachers are expected to cooperate, but I know that in fact the teachers working at the same location never do more than occasionally consult each other and sometimes exchange materials. And yes, I did have a hard time when I discovered this. It really annoyed me. I did try to // I even went so far as to make tests for everyone and hand them out, so that they would all have the same tests and also a scoring and grading system. But then you really get stuck with a lot of work.*

I: Tests they gladly accepted and administered?
[on Dutch "in dank afnemen" : to gladly receive and/or to gladly administer]

J: Yes.

I: Literally.

J: YES (laughter), exactly.

I: And figuratively, yes, yes
(1-1-16)

Joy constructed tests that could be used by her peer teachers as well, but soon realised that the workload was solely on her. Then there was Pete. He differed from Joy's colleagues in many respects. She felt he was an initiator of change in a direction she would like to develop. She warmed up to his ideas and came to adopt some of them. Just as Joy, Pete felt that learners should be able to do more in and with English than simply reproduce items that they had learnt by heart. Pete stimulated Joy to construct her own tests. In these tests the learners had to transfer whatever discrete items they had learned, such as grammar points or particular idioms, to more elaborate test tasks geared at communication.

J: *Well, it happened like this. He (Pete) was teaching upper secondary forms and he observed that, particularly in the havo and atheneum forms, the transition from junior to upper secondary didn't really work out smoothly enough.*
And he felt anyhow that the tests that went with the course materials really weren't any good at all. He even said to me: 'Those kids should really be able to do better.'
Yeah, I mean what good is it when in the end they can only repeat those words, but can't do anything with them, because they themselves can't make that connection, you know.
And he was already pretty far in improving the transition. Every test he had already// for each chapter he had already made his own tests and he told me to do that as well, because I was also in those H, A-forms. I think he also told other colleagues, but if it actually happened I don't know (laughter).
But we did provide a number of things, yes we did.
(1-1-4)

Joy did not simply adopt Pete's views and ideas, as will be shown by the interview section that immediately followed the section above. Joy had already been sensitive to the overall importance of transfer of grammar and idioms in meaningful test tasks.

I: And subsequently you saw those examples. You yourself weren't really pleased {No} with the manner in which testing proceeded. Perhaps that's why you were so sensitive for any//

J: //Yes, and mainly because, as a teacher you were giving it your very best, I can remember, and then I'm referring to those havo and atheneum forms, to teach how to form questions and negations. I think of myself that I was really able to explain that grammar point well and clearly in the classes and the kids really seemed to understand it all when they were doing their exercises and I was walking around, and everything went perfectly well.
And then we had a test with a dialogue in which the idioms had to be used, but when questions had to be formed, nobody thought of using auxiliary verbs. In such situations, they can't make that transfer. So that meant that you really should start preparing them for such tests in class, so that they will be able to make that connection.

I: You really want to see what learners can do in class, they can apply to other situations, such as to the tests they get.

J: Yes, yes, yes.
(1-1-4)

The "//" symbol at the end questions refers to interruptions by the teacher. These interruptions have proved to be indicators of teachers' emotions. Joy interrupts because she wholeheartedly believes in transfer and claims that language tests should contain transfer tasks. She is eager to offer an example of the overall importance of transfer of discrete items to meaningful test tasks. So Joy met Pete at a fortunate stage of her career. During and immediately after the short coffee break we had in the first interview, Joy reflected on the importance of meeting a kindred spirit at the beginning of a teacher's career. The minidisc is recording while we are entering the room.

J: Well, I started thinking about that a bit more because of my colleague, who really seems to have some pretty revolutionary ideas. I don't know how I would have developed, if I hadn't run into him. It's things like that, quite extraordinary really.

I: In fact, they are coincidences that lead you to a certain track you feel comfortable being on.

J: Yes, exactly.

I: And that allows you to develop into a //

J: Yes, because I think that if I had stayed at my first school, I would probably have, yes I don't know, of course, I'm speculating here, but maybe that you would just go with the flow, or something. I don't know.

I: What does that mean for the teaching profession, that you say, well, my personal experiences have been inspired by someone else's ideas' ?

J: Yes, and therefore I'm really a proponent of more discussion and negotiation about these matters. Especially since I have now experienced myself how you can change your views or just how you start to think about certain issues. It's really appealing when you're committed to a certain view, because then you really become motivated to try out your idealistic ideas. And then there are a lot of people who say 'Yes, but you're still young. You haven't been a teacher that long. Your enthusiasm will all end in due course.' But when I look at my colleague for instance, I mean, he might be more of a relativist perhaps, but he is still trying out things and rediscovering and .. . At least that is what I feel.

(1-1-29)

Joy says she is not sure in what direction she would have developed if she had stayed at her first school. What she does stress is the motivating force of cooperating with a peer teacher whose views, concerns and beliefs one comes to share. Joy's cooperation with Pete also influenced her assessment and evaluation practice. At Joy's first school the tests she had to administer were given to her by her colleagues. Almost without exception they were tests that came with the course materials used. Joy calls these tests 'reproductive'. Idioms that were dealt with in class basically had to be translated from Dutch into English in gapfill tests.

J: Well, what I remember best is that first I always used the tests that I got from my colleagues and those were basically the tests that came with the course materials… So actually, yes. And I have to admit that back then I only worked in the junior secondary forms, you know, so those tests were very reproductive.

I: Reproductive you say?

J: Yes, so they asked for a lot of idioms. I can also remember those gap-fill texts that you first read with the students in class and then they had to fill in those words in the test. Yes, so translation really.
Yes and actual exercises from the practice-books that you had dealt with in class and those came back in the tests, as almost exact copies. Frankly, I have to say that most people in junior secondary forms at my current school still work like that, I think.
But at a certain point, I, together with the colleague whom I now work with in the upper forms but who at that time also still worked part-time in the junior forms, agreed that we would construct our own tests.
And he had already started doing that. So he made tests for each chapter and I added to them.
And they really were different, especially for the vwo forms. In the mavo forms you still kind of had it that you really tested particular knowledge and those skills, those words and such. But especially in the havo-atheneum forms,

those gymnasium forms, the focus increasingly was on the application of what they had learnt. So, for example, they got a number of words they had studied and then with those words they had to write a story, or something of the kind.
(1-1-2)

Both at her first and at her present school, Joy's junior secondary colleagues had opted for tests she calls reproductive. The tests mainly consisted of a repetition of exercises similar to the ones the learners had done before. She feels that short-term reproduction does not necessarily lead to long-term retention of the discrete items studied. Joy is also critical of pattern drills and does not feel they are effective.

J: And so what I don't believe in, but then I have already mentioned it ,is that you only //
What they used to do with those drill patterns, where they taught their pupils as many fixed patterns as possible. And where at a certain point they start speaking the language themselves, see, I really don't believe in that.

I: Yes, yes. And based on what experience//

J: My own experience.

I: Yes, yes. That you just see that, yes, you've been doing drills//

J: //and I really can't make any connections with communicative practice.
By the time you go to a foreign country and you have to talk, you're really lost with all your model sentences (laughter). That's just awful.

I: How far back does that go, also to your own education when you examine the way you yourself learned the language? Or didn't drills play any part?

J: Yes, they did in my case. Yes, yes. Clearly they did, in the junior forms.

I: So they also did in your own education. Yes, okay … They didn't have any effect, you say, based on what you can do with a language//

J: No, none whatsoever. The same can be said of grammatical paradigms. Like the ones you had with German. You're having a drink somewhere and then suddenly those repetitions surface in your mind again.

I: Aus, ausser, bei, mit, nach, von, zu (loudly).

J : Exactly, ... like Brigitte Kaandorp ..., but why?
(1-1-31)

In addition to the doubts she expressed about the effects of pattern drills and grammatical paradigms on a learner's communicative competence, Joy expresses her disbelief in the time spent on teaching third-formers everything the teacher had ever wanted to teach about using the passive. (Grammatical knowledge and skills about the passive in English involve transformations and translations such as: They were asking lots of questions. → Lots of questions were being asked.)

J: What we don't really value, by the way, is the active usage of the passive, which they virtually practise to tears in the third forms (laughter). And well I just think that if they know about the passive, that's nice and if they can recognize it, it's nice, but they won't have to be able to use it for the time being. We're quite pleased if they know the difference between the simple past and present.

I: It's interesting that you say we don't really value the construction or use of the passive in conversations and writing assignments. On the other hand, it is practised a lot at school.

J: Yes, but that is (loudly), that is something that is presented in the course materials, you know. In forms one to three they've always used Notting Hill Gate, but now it's Interface, I think.
Yes, and then it's built up like this and in the third form they get to the passive at a certain point and colleagues in the third form find great pleasure in dealing with that subject intensively in class.
And, for sure, it is a challenge of some kind, you know, to explain something as difficult as the passive to your students and then have them nicely reproduce in the tests whatever has been done. But it's rather a shame really. I'd rather have them practise the tenses, in the correct grammatical forms. And then let them write texts with that, or let them do other things, instead of just performing some amazing technical feat.
(1-1-11)

Joy feels the time spent on grammatical technicalities such as active-passive and passive-active transformations is much better spent on having the learners practise verbal tense and form in meaningful written texts. Joy has similar views on teaching and testing idioms and phrases.

Idiom teaching and testing has been a dilemma for Joy practically from the start of her career, when she administered the tests used by the junior secondary colleagues at her first school. Joy objects to rote learning and the subsequent reproduction of a number of selected idioms in isolation. That is why Joy has set herself the task of developing alternatives, in close cooperation with Pete. She wishes to test idioms more productively, e.g. by having the learners do gapfills in authentic texts and integrate idioms in writing assignments and writing tests.

I: Developments. Well, you already identified two clearly important points, didn't you. From reproduction to, well, more than reproduction. And you also pointed out vocabulary as an example of developments that you yourself continuously focus on. Are there any further developments that you could single out, or that you would like to give as examples?

J: No, no. Well, grammar. Last year we decided for the first time to// The foundation, somehow, that was there, but it kind of slipped away you know. At first you just kind of polish it all by simply handing out a few pointers and letting them do some exercises. With sentences where they have to put the verbs in the correct form. But then again, as soon as they get another writing assignment, then, well then they don't remember that present perfect anymore and then even the simple past has kind of slipped from memory again.

I: Comparable to the idiom example.

J: Idiom is kind of the same, yes. And what we started with last year, really from newspaper articles, that was quite a big job. My colleague did most of that, chose articles, which by the way fit right in with the idiom they had studied. So if they had been talking about school, the article would also be related with school. And then we would remove the verbs from the article and the learners would have to put them in the correct forms.

I: So in a way you establish a link in your course materials between the idiom they have studied and grammatical structures.

J: As much as possible, yes. Yes, because that's also something we prefer to do, where we don't just test the language skills, or one single skill, but preferably always have them return as an integrated whole.

I: Yes, with the clear example the way you see idiom and grammar.

J: Yes, yes. But it really is very difficult.
(1-1-10)

Joy prefers integration of e.g. idiomatic and mainstream grammatical knowledge in tests based on authentic texts. Pete initiated the construction of such alternative tests in the year before the year of data collection. Dependent on the type of secondary education the learners attend, the test tasks were such that the learners apply discrete knowledge in meaningful contexts. Thus, *havo*, *atheneum* and *gymnasium* learners were for instance invited to write stories with the words they had learned by heart.

Joy realises it is no easy matter for her learners to transfer discrete grammatical knowledge or idioms to meaningful communicative tasks. She claims most of her learners find it important to do well on tests and therefore prefer discrete items they can learn by heart, such as particular grammar points and idioms. She does not give in, but tries to reason her learners into accepting that rote learning has little long-term effect.

J: Yes, I discuss and reason with them. I'll try to explain to them why we're now working this way. Why that's different than just testing words. That I really appreciate they are good rote learners, but that I also want them to do something with that knowledge.

I: And how do your learners respond?

J: Yes, then they're kind of // Yes, but then, when their average marks barely approach a 5.5, they still keep asking questions like 'Could you just have us translate some idioms, because that is what our junior –secondary teacher used to do?'

I: The perfect way to pull up our grades.

J: Yes, but that's not how I operate. They don't appreciate that. But when you explain and discuss it with them, they do kind of get it. It's just that marks are so important to them. Yes, you really can't get that out of their heads.
(1-1-18)

Idiom teaching and testing is one of Joy's dilemmas. Because her aim is long-term retention, she is looking for effective ways to teach and test vocabulary acquisition. Her plans for the present school year is to get rid of discrete idiom testing altogether.

J: And yes, the way in which you test idioms is always problematic, unless you simply say just study this and just translate that. But also in these instances, I would write my own texts or find authentic materials. But now I think we're at a stage where we will radically change our approach to idiom teaching and testing. We have them study vocabulary or have them compile word files from existing texts, which they have to revise, keep up and study. Yes, and then we try to take these files into consideration in the tests we make. For example, in writing assignments you attempt to have them use vocabulary from their word files.
(1-1-6)

Joy plans to make her learners more responsible for the idioms and phrases they have to retain by having them produce individualised word files, instead of testing them for vocabulary at regular intervals. It was a decision she would come to regret towards the end of the school year. In the first interview, we followed up Joy's preference for integration of knowledge and skills and authenticity.

Because of her orientation on transfer tasks, Joy started to develop a preference for integrative testing by using authentic texts whenever possible. In these integrative tests discrete knowledge and skills of various kinds have been integrated. An example is a test task in which the learners have to write a meaningful story and use an array of idioms and phrases that had resulted from their rote learning. To enable transfer, Joy feels that a meaningful test should be based on authentic texts. If authentic texts proved to be a little too difficult for some junior secondary classes, she would provide a meaningful context herself.

> *J: I would write the words they had studied in a little text and then they// or I found a text somewhere. But the latter did prove to be quite difficult in the junior forms. If you want to use authentic material, it's often simply too difficult, so usually I would write texts myself. And then I didn't have them translate the words, but have them insert the words in the right ways, in the correct places. And in the course of the year, I'm referring to a first or second form, then the learners also had to provide the correct grammatical form. (Yes, yes) So each time we progressed a bit.*
>
> (1-1-6)

We were curious how Joy would assess that a particular authentic text would be too difficult for her learners to understand. This is how she responded.

> *J: Because I did try it a few times and then you always come across idioms they just haven't yet mastered. And that's really difficult when you// even in children's books, English children's books, then even those prove to be just a bit too difficult for that approach of testing. So you really have to write a text where you know they really should be able to understand all the words.*
>
> *I: So it's important to you, there really shouldn't be any idiomatic problems in your texts such as the application of questions and negations.*
>
> *J: Exactly, exactly. Because then they panic and then uh.*
>
> (1-1-7)

So finding texts of the appropriate level is basically a matter of trial and error. When Joy uses authentic texts, she is aware that they might endanger the validity of the test. She is convinced that a test should assess what one intends to assess and feels the result of test scores should not be affected by a lack of idioms the learners cannot yet be expected to master. The more advanced the learners become, the easier it is to use authentic texts, both as test input and as part of the test output, e.g. by having the learners fill in or change items in an authentic text. She also claims that her preference for integrative testing can be problematic as well. Joy says they are hard to score in valid and reliable ways. This particularly concerns writing tests.

> *J: Yes, what we think a test should have is good criteria, that is to say the demands. But that's really quite difficult. Especially with writing assignments, scoring and grading is kind of loose and pliable and it will remain that way.*
>
> *I: Speaking of criteria, you mean criteria for scoring and grading?*

J: Yes, yes, what you focus on, yes. We really think they should be clear and that we should inform our learners about them beforehand: 'Well, this and that is what we're focusing on.'
And at the same time, what we would do with such a writing assignment, if I don't feel very confident or something of the kind, I would have the test checked by a colleague, and then we each grade the assignment individually. Or I would have a look at what he has done. Of course, there are always certain things you focus on. Grammar should meet a certain standard, you focus on word order, the spelling a bit and you focus on content. But, yes, even then each individual case is so different.

I: There is still a whole grey area (yes) for which you would like to have some criteria//

J: //Yes, but this even goes for assessment criteria for spelling and grammar. How do you score and grade a certain mistake? And, yes, what's really annoying is that there are all these nice grading sheets, but I would really like to develop them myself, so that really is very difficult (laughter).

I: Yes, definitely. So you're saying that when you look at a grading sheet you think "Hmm, do I really agree with this"?

J: Yes, exactly (softly).
(1-1-23)

Amidst her concerns for reliability, validity and fairness to the test takers, Joy acknowledges that constructing effective scoring and grading criteria is a formidable task. She feels this is even true for areas that are expected to be clear, such as spelling and grammar. We will further illustrate and explore Joy's concerns and beliefs, when we discuss the three sample tests she was to select in the course of the school year.

In telling us about her teaching and testing experiences, Joy regularly referred to a link between learner autonomy and language learning motivation. The interviewer first summarises what she had previously told him.

I: Regarding learner autonomy, you earlier pointed out in our interview that it is very important to learners that when they do something, they're motivated. We talked about the things that learners enjoy and also about aspects that you found easy. It's easier to just lean back and be told what you have to do, than to be challenged and to be active yourself. That's just horribly//

J: Of course each learner is different. In the case of English, not every learner is going to like the subject. in a way they kind of see the importance of English for the future, but they nevertheless don't really enjoy it. They may, for instance, be more interested in technical classes, for instance.
And then literature is even more of a bore, in general. Yes, then a learner would rather just have you tell him what to do, so that he doesn't have to put a lot of energy into it. The curious thing is, though, that the more freedom they get and the more they're allowed to think about what to do with it, the more they will regard it as something altogether positive. That's true.

I: Almost like two sides of the motivation quarter.

J: Yes, it's very ambiguous.

I: On the one hand just tell me what to do and on the other well this is what I like to do.
Extrinsic and intrinsic motivation.

J: And I can't really point out when there's talk of the one or the other.

I: No, no. I would find that very difficult as well. Well, I think that when we re-read the interview, apparently there's an aspect of motivation in learner autonomy. You already talked about the dialogue with your class. If that still plays a part in your assessment and evaluation practice, we could consider it more closely.

J: Yes, I think motivation plays a very big part even.
(1-1-34)

Joy's conviction of the strong link between learner autonomy and motivation was the last belief we deducted from the first interview. Learning motivation and teaching motivation were often assigned as codes to segments of the interviews we had with all of the three respondent teachers. That is why the links between language learning, learner autonomy and motivation will be discussed in more detail in chapter 12.

Looking back on Joy's formative experiences and core beliefs, we see that Joy has referred to three main formative experiences:

- The insight and understanding in how English works when she actually started explaining matters to a relative of hers;
- Her experiences at her first school, where:
  - teachers did not cooperate much;
  - uncritically stuck to the course materials that were being used;
  - teachers did not develop materials of their own;
  - she discovered that it was hard for learners to transfer discrete grammatical or idiomatic knowledge to communicative tasks;
  - she developed an interest in transfer and a preference for using authentic texts and tasks in teaching and testing.
- Her experiences at her second school:
  - where she noticed differences in views concerning effective teaching and testing between her colleagues working in junior secondary classes and her colleagues working in upper secondary classes;
  - where she was disappointed at what the young learners were expected to know about and do in English;
  - where she met a kindred spirit in the person of Pete, which helped her discuss and develop the initial ideas she had and who was a catalyst in producing materials and tests of her own that focused on meaningful communication.

Towards the end of the first interview, the teachers were asked to define the constructs that were central to the investigation. We will report on these construct definitions in three sections. First, we will address how Joy interpreted *learner autonomy*. The construct was presented in the jargon common in the Netherlands at the time of data collection: 'self-regulated learning' (*zelfstandig leren)* and 'learning with full learner responsibility' (*zelfverantwoordelijk leren)*. Next, Joy defined the supposed be-and-end-all of foreign language education, a construct often used but hardly ever well-expressed, i.e. *communicative language education*. Finally, we will present what Joy felt to be effective *language assessment and evaluation*.

In the next section, we will see how Joy interpreted and defined learner autonomy.

7.3.2 Learner Autonomy

Joy claims she feels fairly comfortable with the construct of learner autonomy. She defines how she views the concepts *self-regulated learning* and *learning with full learner responsibility* without hesitation. The reason she gives is that she has recently finished her degree-one teacher training. Here, the focus has very much been on the theory and practice of learner autonomy. Besides a fundamental discussion of theoretical principles of learner autonomy, the trainees themselves had experienced, assessed and evaluated these principles in didactic procedures both at the institute and at their schools[2].

> *J: Yes. Yes, I have been able to think about that some more over time. I think it's very important to teach learners how to think more consciously about what they're actually doing and what they should do, what they want to do. And how they could accomplish that.*
>
> *I: You want to involve the learners by asking questions such as "what does learning actually mean to you?"*
>
> *J: Yes, yes. And by taking baby steps, because, well, because it appeared that is what they find quite difficult. And it is difficult.*
>
> *I: Yes. You're saying you've thought about it more in the meantime. For the sake of this study, that requires an explanation. You say "I've started to fill in that concept." How did that come about? That you started to fill in the concept of learning on your own according to a working definition, saying that: To me, learning on your own means….*
>
> *J: Well, because here at UNILO I noticed at a certain point that with self-directed learning and full and effective autonomy, I still ran into problems, one way or the other. And, as far as the learners are concerned, especially in the final forms, they still appear to be inclined to lean back and say 'well, come on and tell us what to do.'*
>
> *I: I see. So learners expect that you, the teacher, {Yes}, tells them that this and that {Yes} should be studied and then we just go on and do that//*
>
> *J: Yes, exactly, and parents have that tendency too. They feel that the teacher is actually the person who is responsible for their children's results. Learners themselves also think like that sometimes.*
> *And now I've noticed that if you take the time to discuss matters for a while, they feel autonomy makes sense and start thinking about what they would actually like.*
>
> (1-1-32)

Joy feels that the basis for self-regulated learning is that learners consciously attempt to reflect on what they are actually doing in or with English, on what is expected of them, and on what they would like to do themselves. She adds that the learners also need to think about the ways in which the national or school targets as well as their own targets can be realized.

She admits that this is a formidable task for the learners, used as they are of sitting back and carry out tasks they are asked to do. That is why Joy prefers a gradual transition of teacher control to more learner control. She feels that both the learners and their parents must realise that a teacher is not the only person responsible for the teaching and learning process.

In view of the concept of *learning with full learner responsibility*, Joy feels this is too far removed from reality. She is skeptical about its feasibility in school or classroom contexts.

J: Well, learning with full learner responsibility, I don't think you can accomplish that in secondary school. I don't think you can do that. Actually, you have to, yes, you're really bound to those attainment targets.

I: How would you define learning with learner responsibility?

J: Erm, well, it's actually what I'm doing right now. At a certain point I notice that I'm not doing enough with my speaking or conversational exercises and then at that point I feel I have to do something about it, which I then do.

I: Choices evolve within the learner and the learner decides whether to act on them or not? (Joy nods in agreement) I agree.

J: And up to a certain point I can, but always within those attainment targets, you know. There are learners who are saying well, I don't know, especially towards the end of the fifth or sixth form when those last weeks are optional. Then the learners can decide for themselves what they still want to work on, right, so they have to set their own goals. Yes, well I'm not sure if that actually is learning with learner responsibility.
(1-1-38)

Joy feels that are too many restrictions and constraints that prevent the learners from taking and accepting full responsibility for any action they undertake. She specifically mentions the targets that have to be attained in the course of their education and feels that control over her learning as she, as a beginning teacher, has, is out of the reach of her upper secondary learners.

7.3.3 Communicative Language Education

Despite Joy's preference for transference of discrete knowledge and skills to meaningful communicative tasks, she has difficulty in defining what communicative language education involves. Why this is so is caused by the fact that she finds it hard to determine what the roles of grammar and vocabulary are when learners learn to communicate in English. Without some basic knowledge and skills or basic sense of how a language works, Joy feels learners are likely to become 'unguided missiles'.

J: Yes, I do find communicative language education quite difficult.

I: Yes, it is a difficult concept indeed.

J: I do think that you// because of course there's been this whole discussion about what "communicative" exactly means within basic secondary education, you know. Do you really chuck out all the grammar and just, well, just teach learners to express themselves clearly enough in that language? I don't think so. I think there has to be a good foundation, because otherwise, well, sometimes it seems like they just become a kind of unguided missiles (laughter). Or they just don't have a clue how a language works. And maybe that has something to do with gaining insight in, you just have to have certain// you have to gain some insight into certain structures and the way in which a language works. But the difference then is that I, for example, when you test, I mean that foundation has to be there, but if I'm purely testing their communication then I find it quite difficult to say, well, this is a grammatical error, but is it communicative or not. I really find that hard to do.

I: Does it have anything to do with criteria? Is it actually possible to define clear criteria//

J: //yes exactly.

I: and to let something pass or not.

J: Yes exactly. And then it's almost always a question of just feeling your way and well if you were in England and you would say or write this, then I think they would understand you (grins). But, yes, well.

I: So once again the foundation is important, that you don't become an unguided missile and//

J: // I agree. Yes, I really think that even at the vbo-level you have to explain at least once that now, listen, if you ask a question in English, this is the difference with Dutch. And neither deal with it too deeply, nor talk about it for too long or anything like that. You just have them practice it once and then you hope that at a certain point they think 'Right, it goes like that.'

I: And here you make the difference very explicit, where you say that you demand something else from someone at the vwo-level, or you have different expectations than say from someone at the havo-level or a//?

J: Erm, not from someone at the havo-level, but from someone at the vbo-level//

I: Yes, alright. Well that difference//

J: With vbo-learners I'm easily satisfied.

I: Yes, yes. But differences between havo-learners and vwo-learners, what do you//

J: No, no, not in junior secondary forms, hardly any, no.

I: So what you're saying about communicative language education, well, it is about applying what you can do with a language in practice, isn't it? So for communication that means the processing of information, or the production, doesn't it? But to be able to do so, you need a foundation.

J: Yes, yes. So you really have to pay some attention to grammar and you have to study some words to build up your vocabulary because otherwise erm//

I: Yes, all right//

J: // And, yes, you occasionally have to test this as well.
(1-1-30)

Joy's problems to define what communicative language education is, seems caused by her difficulty in determining the roles of grammar and vocabulary when learners learn to communicate in English. She feels some knowledge is needed, but finds it hard to define exactly what. She also finds it hard to determine in how far a grammatical error is in the way of effective communication. This means that as a teacher Joy relies on her feelings and intuition, rather than on criteria that indicate whether a particular utterance is communicative or not. This dilemma is not exactly solved by colleagues, who appear to hold different views on the roles of grammar and vocabulary in teaching and learning English. All in all, Joy seems to find it easier to define what communicative language education is not. She is quite convinced that drills or a complete focus on discrete grammar are of no help. Neither does the reproduction of idioms or useful phrases you have once learned by heart. I ask her about both the clarity and the practicality of the concept of communicative language education. This is how she responds.

J: Erm … Well it isn't very clear at all, but I do think it's workable.

I: I see. So what's missing? What isn't clear about the concept of communicative language education?

J: Well, yes I mean that everyone//What I mean is that in a way everybody has their own opinion about what is and what isn't communicative. It's very difficult to see eye to eye on it, even amongst colleagues, to bring it into line, or to come to an agreement. Yes, and when you look at those basic secondary education tests. When I had first administered them, I thought 'Well, what am I so worried about?' They will all be able to pass. I thought, well thank goodness (laughter). But//

I: // You teach them to communicate, so the problem's solved?

J: Yes, exactly. But having said that you would still like a bit more clarity, you know, towards each other.

I: I get your point.

J: Because what you're hearing now is that in this particular with colleague so or so, well, no grammar is done, for instance. And with the others they pay a lot of attention to structure, and yet another colleague takes up a middle position etcetera.

(1-1-31)

Communicative language education is a feasible concept for Joy in the sense that she manages to teach her learners to communicate in English, as shown by the results of her learners on the tests of basic secondary education. However, the concept appears to be far from clear to her and some of her colleagues. In chapters 11 and 12 we will elaborate on the clarity and practicality of conceptions of communicative language education, with its construct of communicative competence.

After Joy's interpretations of learner autonomy and communicative language education, we will now present how she views effective foreign language assessment and evaluation.

7.3.4 Foreign Language Assessment and Evaluation

In the first interview, the teacher respondents were asked to define essential knowledge, skills and understandings in English they felt were important to test in the course of a school year. The respondents also defined qualities of what they considered to an effective English written test. Finally, they were asked to present their views on the potential washback of tests when learners learn to communicate in English.

Essential knowledge, skills and understandings

Joy claims that basic linguistic knowledge is regularly part of the tests she designs and administers. She mentions four structural aspects (1-1-11):

- *ability to formulate questions and negations "in reasonably correct ways";*
- *ability to put verbs into appropriate tenses and/or forms;*
- *knowledge of the declension patterns of the irregular verbs;*
- *correct word order.*

Next to informing us about basic knowledge and skills, the three respondents defined how they viewed the term *insight*. The question had been included because objectives in the Netherlands used to be formulated in terms of knowledge, skills and

insight. Joy found it hard to specify what the term includes. She feels the effects of insight and understanding are there, but it is hard to describe or pinpoint them.

Joy gives three examples in which insight plays a role. The first example is related to learners scoring well above their average marks for English when they prepare for and do the national CITO reading comprehension examinations that make up 50% of the final mark a graduate gets for foreign languages.

> *I: Does insight play a part in learning to communicate in English?*
>
> *J: Yes, well, do you mean that some learners may by nature be more insightful than others?*
>
> *I: For example, for example.*
>
> *J: Yes, that does play a part. Yes, it definitely does play a part.*
>
> *I: And can you relate that fact to assessment and evaluation, or do you feel that is difficult?*
>
> *J: Well, that's something// Well it can be, well not frustrating, of course I'm very glad when learners have insight and they're good at English, but, for example, there are also learners who have insight in understanding texts. They're very good, just like that. And what I really find quite difficult, they really think they're very good, get high grades, what I really find difficult then is to motivate them to really explore texts. You usually accomplish that by just discarding the multiple choice aspect completely and then really start working with open questions. And then it is a shame that those learners who really work hard have less insight. They generally get lower marks, and to me it is a challenge to see how they go about it and see what we can do about it. But I have to honestly admit that at the moment of testing we don't take that into consideration much. No.*
>
> (1-1-25)

A second example in which insight and understanding allegedly play a role of importance is the case of one of Joy's learners who appeared to be an exceptionally gifted writer in English.

> *J: And then you also have learners, now I'm talking about this boy, who is a fantastic writer, and who also has an enormous vocabulary. I don't know from where he has got it, because I'd always thought that maybe he'd at an international school or maybe he'd lived somewhere else but. It seems to be the case that he writes just as beautifully in all languages (laughter)//*
>
> *I: // Somehow. Like it just comes naturally.*
>
> *J: Yes, but in any case, I don't adapt my tests to him as an individual. My tests are geared to the average learners, and then he scores nine out of a maximum score of ten, or something of the kind. What I do try then is afterwards to talk about it with him, so that he can tell me a little more or something of the kind.*
>
> *I: What do you mean by talking to him some more, by letting him tell you some more about it?*
>
> *J: Yes, well, that he can, that I let him know, well, my goodness, this is great stuff. Tell me how did you, because usually it's about an essay or something, how did you get this idea, or thought, if it's about a review.*
>
> *I: Insight is difficult to define and lacks transparency.*

J: Yes, I do feel the same.
(1-1-26)

The third example Joy presented has already been presented as a first quote in this chapter (1-1-27). It was the moment when Joy became more insightful herself, when she taught a cousin of hers about the language. It has already been discussed as one of the main formative experiences.

Joy was also asked how she would define an effective written English test.

An effective written English language test

Joy first mentions two aspects of an effective test. She says that tests have to represent the level the learners are at, and should always be a little more difficult than what can reasonably be expected of the learners. Her view echoes Krashen's N+1 principle, which she applies to language tests. Second, Joy says tests have be transfer-oriented, i.e. test tasks must be meaningful and geared to improving communicative competence.

J: Well, I, first of all I think it's important that they are at the level where the learners should be. I try to// They should always be just a bit more difficult than usual. But not all the time. It kind of depends.
But, it should fit in with what they can do and what they know. And next, they should always be transfer-oriented.

I: Right, yes. Maybe we could go into that first bit a little bit deeper, OK. Where you say that the level should be at least equal to//

J: //yes, plus one or something//

I: //or plus one, where you make it a bit more challenging. Why is that important to you?

J: Well also because of the incentive. I think that it challenges learners to work towards the test more seriously than they would do otherwise. I think it kind of works.

I: Do you have an example in mind where you think 'well this and that happened in class and I would like to spice it up in the test a bit by asking it in this and that manner?'

J: Erm … (4s) Let me think… Yes, then I have to come back once more to// it probably seems like all I test is grammar and idiom, but that really is not the case. (laughter). But that's the way it is, I suppose. (loud)

I: You referred to them as foundations, of course, and then, yes.

J: Yes, precisely and you have them practise grammatical structures a bit and you have them fill in authentic texts a bit, the way they will also appear in the test. And of course we don't give them detailed information, but we inform them nevertheless, because they should know what to expect. But then again we always choose a text that's just a bit more challenging. Or the texts they practice with become a bit more challenging each time and they get more demanding each time.

I: The structures become more complex.

J: Yes, yes.

I: I agree. Of course it's very difficult to talk about tests without //

J: Yes, without//

I: Without actually seeing a test in front of you, isn't it?

J: Yes, very much so. (laughter)
(1-1-19)

Joy then gives an example of a language test she felt was effective.

J: Well, I was very pleased that we had done a lot of school idioms and that we had also read a lot of texts. They had also looked up their words independently and synonyms and what not. And then at a certain point I linked it to a writing assignment in which they had to say something about themselves. So that was after the incident when after two months I had found out that they didn't know anything anymore. So then I also adapted my test, and// Well, so they had to take on the role of the teacher and I was the student and they had to report to their parents what my improvements in certain areas were.

I: In English?

J: Yes, in English.

I: And in which form was this?

J: Erm, the fourth form.

I: Ah, also a fourth form.

J: Four havo, four atheneum. They really enjoyed doing it (laughter). And I left them quite free to establish whether I was doing well or not. But they did have to meet certain criteria, had to do it about a few school subjects, and, yes, the idioms that had been studied had to be used.
(1-1-20)

After thinking for a while, Joy offers a second example. This test is comparable with the third test we were to discuss in the course of the year.

J: Well, kind of a combination of listening, writing. Also a bit of literature. As a project we dealt with homelessness and the movie adaptation of the novel Stone Cold, by Robert Swindells. So, they watched the movie in three episodes and they had to take notes.

I: OK, yes.

J: And through the strategy of note-taking, they had to summarize the book and subsequently they had to play the part of that boy or that girl who was a journalist, and//. Or that journalist could write a piece about what they had done and that boy who wanted to write a letter to her and who felt cheated, and so on and so on.
(1-1-21)

The test strongly hints at a didactic preference Joy has come to share with Pete, that is. an interest in project education in which knowledge and skills are integrated and presented in a variety of challenging tasks. We will learn more of this preference, when Pete, the project man, has the floor.

A final aspect of language assessment and evaluation we discussed were Joy's views on washback.

Interpretations of washback

Joy feels that tests are not the instruments to help learners learn how to communicate in English. The actual learning takes place in the period preceding the test. She nevertheless acknowledges the importance of test marks for her learners. However, Joy does not actively use the test and its assessment criteria to guide learning and determine what the learners learn, how they go about it, and why they do so in the chosen ways. She does not refer to the role of test evaluation in the first interview. Joy primarily sees tests as summative assessments, i.e. as a tool "to check whether a learner achieves what he is expected to achieve" (1-1-39) at the end of a learning process.

> *I: What role do your test items and the tests themselves play in the process of learning how to communicate? So now I'm talking about the tests that are scored and graded. What is the role//*
>
> *J: Yes, I// Learning how to communicate already starts in your lessons, of course. And a test like that is really just a means of seeing whether a learner is sufficiently capable of doing that. But actually, the way I feel about giving marks, it's just that they kind of motivate the learners. Focus on achievement, in a way that's the way society works. So because of that you have to work towards that, but on the other hand, well, you also just want to see whether they can achieve a certain level, whether they have achieved that level.*
> (1-1-39)

Joy feels that marks are important for learners. Electronic calculators spontaneously arrive on the tables, when test results are made known. Joy senses a difference here between the average havo-student and vwo-student, the former being more calculating in more than the literal sense, and the latter being more ambitious.

> *I: You've mentioned that, of course marks are not your primary concern in the end. You acknowledge their importance, that they have a certain effect, but the main thing for you is to clearly see what your learners know and what they can apply. What's the learner's role in this? How important are marks to your learners?*
>
> *J: Well, very important. Yes, goodness me, they are educated that way, aren't they. I mean they find them very important and you really notice that the more they approach atheneum or gymnasium-level, the more important become even tenths of marks (laughter). Then it's really like, those calculators appear whenever the slightest need arises, and they start calculating. Yes, that's erm//*
>
> *I: //so between those levels you see differences, whereas a havo-level learner might just shrug his shoulders once and think well, better luck next time.*
>
> *J: Yes and you also really notice it with re-sits. If a re-sit is possible, just to find out whether you can prepare yourself better the second time. If a havo-level learner gets a 5,5 and he calculates that his mark is a 5,6 average, then that's okay to him. Then he's definitely not going to re-sit the test. He feels that is quite OK. An atheneum-level learner will try to at least get it up to a six. That's just the way it is, even though when you look at the difference in the level achieved in the end, it's just barely there. A few tenths of a mark.*
> (1-1-41)

Yet, Joy is of opinion that the potential power of language tests to affect the learning should not be overrated. She thinks it is a fallacy to believe that learners only work when there is a test ahead. In the course of the interviews Joy had given examples of learner tasks that had proved to be motivating. She feels such tasks are more likely to affect the learning process than the actual test at the end of a learning cycle. Nevertheless, a lot of her colleagues believe that administering lots of informal tests ensures that the learners learn what they are expected to learn. The pressure of marks is essential to many of her colleagues. Unfortunately, Joy feels that the battery of tests her colleagues envisage are basically on testing discrete knowledge or microskills. Joy's colleagues generally do not believe in diagnostic tests, because marks are missing.

> *I: So how important do you think tests are, within the PTA in the second phase?*
>
> *J: …Erm…yes they // everyone has a different opinion about it, I've already noticed. Some people really stress the importance of tests, because they think, or they feel that learners only start to work because of those tests, if you know what I mean. So the tests and the marks learners get are seen as the only stimuli for learners to get working.*
>
> *I: Which basically results in lots of tests, because at least then learners are expected to work//*
>
> *J: //lots and lots of tests. And then preferably a lot of microskills are going to be tested. That's not allowed any longer, but people still want as many tests as possible. And diagnostic tests and self-assessment. Well, people are really quite sceptical towards that. People still think that the learners, well, that they think, it isn't registered anyway, so I'm not going to give it my best.*
> (1-1-42)

The first interviews with the teachers ended with their plans and intentions regarding teaching and testing for the school year to come.

7.3.5 Joy's plans and intentions for the school year

When Joy is asked to tell me about any plans and intentions for the school year that is about to start, she mentions experiments with self-assessment and peer evaluation by the learners. Joy refers to a learner and teacher resource book ("Bronnenboek", Mulder 1998) that contains tables and schemes that can help learners to assess their products and learning processes themselves. However, the plans have not reached the operational stage yet.

> *J: Yes, the "Bronnenboek," (Resource book for independent learning) so then the learners can finally assess themselves. We're going to start to experiment with that this year as well: what they should focus on, the way a text is constructed, but then again, also the grammar//*
>
> *I: Yes, and then you're talking about self-assessment in general or focused on a certain//*
>
> *J: No, in general.*
>
> *I: Yes, alright. So you're saying that, well, you didn't mention it as part of an effective written test, but that an element of self-assessment is also required in it. But you do make that link between the right criteria and self-assessment?*

J: Yes, yes. I do believe we're also definitely going to start experimenting with that this year//

I: And so it's also a resolution that ehm//

J: Yes, yes. (1-1-??)

Diagnostic tests have been included in the PTA (Programme for graduation and summative assessment). The example Joy gives concerns the assessment of writing skills. The levels and the demands of the learners will gradually be increased up to the final examination of writing at the end of the school year.

I: Did you include those in the PTA for English?

J: Diagnostic tests?

I: Yes.

J: Well, yes, they are mentioned yes.
For example, in that first writing assignment, they're.
Right before that they get a diagnostic test and uh, they'll take the test and then they'll test each other in pairs. And because this is the first time, we'll take their tests back and then discuss the results with them.
Well, yes, then you've evaluated one another in this and that way, but you could also check this and that.
Deep sigh

I: Do the criteria still play a part in that? At the time that//

J: Yes… that they test each other? Yes, I do believe that that's going to be very difficult.

I: But you didn't make a fundamental choice where you said well we'll just leave that open or do we actually define a certain structure already?

J: Well that kind of depends on the test, of course. But regarding the written test, erm. Yes, in any case the foundation should be solid and they should be able to build up the text in a logical fashion, so that it doesn't become a, uh, collection of loose scraps or something. And then each time you take it a step further. Maybe that if we mainly focus on grammar in the first test and then the second test, ehm, well then we'll also focus more on the content, the build-up, and then in the third test the grammar should be correct and, the build-up should make sense and they should also be able to say something about the subject they're writing about.

I: Yes. Yes, that sounds very interesting of course. Could you predict, based on this, whether you're going to start working in this fashion or whether it's still so new that you really //

J: //no, it's not really new. I don't know… It's still a bit, experimental …
(1-1-23)

Peer assessment will be a focus for the year to come, but the plans had not really crystallised at the time of the first interview. Besides, Joy predicts that the construction of assessment criteria for peer assessment will be very difficult.

Joy finally highlights some important components of the PTAs she and Pete have produced for 4 havo and 4 atheneum, i.e partial use of course materials (*Touchdown*) that has been designed to have learners partly work and learn on their own, an

integrated approach of literature teaching, the demand on the learners to file aspects of the integrated tasks they have performed in the files that have been named after the traditional skills of *reading*, *writing*, *listening* and *speaking*.

I: Yes, in fact you've already answered part of the last question, about your decisions and resolutions. For example, the phasing such as occurs with the written assignments, within the PTA, now that's quite characteristic of the second phase approach. Are there any other concrete decisions or resolutions?
That you would like to bring into practice next year?

J: Well, … Not really, I don't think so. I'm just thinking whether we've done anything else.
Yes, well, what they really have to do entirely on their own each time is that 'Touchdown' thing, that text book where they have to find out how to construct certain pieces of text and uh, also with writing and speaking. So they really have to work on their own with that.

I: So they have to make use of the grammar they can look up and then use and apply that.

J: Yes, yes, it's totally new for us too.

I: And, of course, there is the compilation of word files.

J: Yes, yes, exactly. And what's also new for us is that we're actually going to use a course book.
You can find that back in there sometime, but then it's also to let them know that now you're going to use 'Touchdown', look, it says so right here, this is a writing assignment eleven point two and you're going to work completely on your own with this. We really don't come into the picture anymore then.

I: And alone or in groups or //

J: // in groups. Pairs.

I: All right, in groups … Well, that's quite something, all the plans you have for the year to come.

J: What's also a bit new is the literature part, you know. We haven't totally agreed on that yet.
In principal it will be part of integrated literature education , but somehow it didn't quite work out, so now I've made a design about how and what, together with a Dutch colleague.What we decided is that the literature of all the modern foreign languages will be placed in the literature file for Dutch and that he manages that and that happens according to certain rules and then we check things off the list, so that he retains an overview of what's happened.

I: And most of these assignments will also be done in Dutch, of course? Or does English still play a part?

J: Ehm, yes well I think that we will just combine the two, but how that will all turn out exactly I don't know yet. Because they still also have a file for English, you know. What they get for writing, and also for reading, but then focused more on multiple choice and such. Erm, listening, speaking file and things like that. It's just the literary component that goes in the file for Dutch.

I: So now you're saying, well, yes, the integration of skills is in fact important but the files that are being compiled have already //

J: // yes that's quite strange, we've already discussed that because of course it's quite strange.
It's just that it's also very new to us. That in the lessons we actually, Pete also brought that up, yes that's very ironic. That in our lessons we integrate the skills, but then we make our learners split them apart into, well, just look what goes in which chapter of your file. That's quite curious actually.
So I'm interested to see whether it will work or whether we'll just go back to, well, how to do it, yes, there'll be lots of questions about that. Does that fall under reading or somewhere else?

I: Well, I'm very interested as well.

J: Yes, I am too actually.
(1-1-44)

This is where the first interview ended at the beginning of the school year. Even though Joy is a capable planner and organiser, there will still quite a few uncertainties at the start of the school year. The task for Joy, and all of her colleagues working in the Second Phase, seemed formidable at the beginning of the school year.

7.3.6 Joy's language tests

In the course of the year we monitored Joy and her assessment and evaluation practice in her 4 atheneum class. Throughout the year we discussed three written tests in detail. Joy was free to choose a written test of her preference, as long as she felt the test was representative of her approach to testing and that the tests contained knowledge and/or skills she considered relevant to be assessed. We will briefly refer to the tests she was to select in the course of the school year.

The first test Joy selected was a writing test that closed off a project called *Dear Nobody*. The project was based on Berlie Doherty's novel on teenage pregnancy of that name. The learners are asked to carry out one of the three writing tasks presented to them.

The second test Joy selected was a gapfill based on an authentic text. The learners had to put 30 given infinitives into appropriate tenses and forms. The authentic text the test was based was entitled "Little Boy growing Old before his Time".

The final test Joy was interviewed on was a reading comprehension test. The test was based on an article taken from *Time, July 12, 1999* on the positions of Tony Blair, David Trimble and Gerry Adams in relation to the Good Friday Agreement. The English text was accompanied by twelve open-ended questions that were presented in Dutch and had to be answered in the Dutch language as well.

Without any suggestion or pressure from the interviewer, Pete opted for discussing the same three tests that Joy had selected. It allowed for interesting comparisons between Joy's and Pete's responses to the same tests. Yet, before turning to Pete, we will first concentrate on Mark. His love of English and American literature and his criticisms of *basic secondary education* and the *second phase* prevailed in the interviews we had with him and were illustrated in the language tests he was to select.

7.4 Mark, the literary master

Mark boasts twenty-three years of teaching and testing practice when we have our first interview. He is the only respondent to make a comment while reading the first interview guide. "Linguistic skills, not literature", he utters with deliberation,

shortly after he started reading the interview guide. The utterance later appeared to be an early illustration of his sincere belief in the opportunities of literary texts when English is taught and learned.

Mark is precise and detailed in reading the guide and takes down some notes. Having read the guide, he warns me to tell him straightaway when he raises issues that seem to be outside the scope of the first interview. Mark is eager to get things off his chest. We never bother to stop him. In free-attitude interviews it is essential for respondents to be able to express their main concerns with as little intervention as possible. Instead, we encourage him to say what he wants to say. This way, Mark's core beliefs in teaching and testing become clear from the start.

We will report on the results of the first interview we had with Mark similar to the way in which we discussed Joy. First, Mark's formative experiences and core beliefs will be focussed on. This will again be followed by three sections on how Mark defined and interpreted learner autonomy, communicative language education and foreign language education. We will end our discussion of Mark with his plans and intentions for the school year to come and a brief introduction to the language tests he was to select in the course of the year of data collection.

7.4.1 Formative experiences and core beliefs

Mark started his career as a junior secondary teacher at his present grammar school in 1976. He had just got his MA in English language and literature at Nijmegen University. Mark had done his thesis on American Literature. His academic training has proved to be formative in the sense that it had developed and shaped his love of literature and the didactic approach to literature he was to adopt. Academic education in Mark's days was predominantly characterised by the academic educator being in control. Academic assessment involved writing essays on novels and poetry, and courses were assessed by way of oral examinations. All of these elements recur in Mark's teaching and testing practice. This is how Mark looks back on his role as a teacher at the beginning of his career.

> *Twenty years ago you were indeed at the classroom centre and ex-cathedra teaching was the way to do it. .... You starred in your own show. You still do that, to some extent. But even then, in the sixth form, I approached literature classes in such a way that I started each class with a short story. Each week they would read a short story from a collection of stories I think they also used, or even still use, at the university. Starting with Young Goodman Brown and then Bartleby. And then I'd give a short introduction or some information about the author. First, we would discuss the story in a circle, and this was all done in English. The learners would take notes, in a kind of academic way. Then, after I had done that for about two or three times, each week thereafter two learners would prepare the lesson, preside over the classroom discussion, and I would just pull up a chair.*
> (1-2-5)

The segment illustrates Mark's preference for the materials and didactic approach he had experienced as a student in the world of academe. The short stories he presented to his upper secondary learners were the ones that had been discussed in academic workshops he had attended himself. The didactic approach Mark initially adopted reflects the way in which literary texts were discussed in his university days. The educator sets the example and determines the level and depth of the literary discussion, which is later to be copied by the students themselves.

His experiences as a student with assessment had also influenced his assessment and evaluation practice as a teacher. Mark has a preference for having his learners write literary essays as formative assessments and by finally assessing them in oral examinations at the end of a school year.

Mark claims his didactic approach to literature has changed over the past few years. Mark expresses his regret that his grammar school learners only have to read three English novels at the end of their school career.

> *M: Yes, I would regret that very much, but then, once again, I am speaking as a literary man and in fact all of us in the department are literary people, and also on the basis of what you could call the success over the past years. Of course we use what I, I don't know if you've got that on tape, but that which I mentioned we used in the beginning. You know, the academic method.*
>
> *I: The example of the circle in which you...*
>
> *M: Yes, but we let that go a long time ago. Now we are working much more on a project by project basis and have them work in groups and that has more to do with the didactic procedures, the independence, the skills involved in doing your own research, the teacher no longer being the sole source of information, and learning to cooperate. I mean, teaching learners how to evaluate literary texts, in short, we've really been focusing on that for years now. First by experimenting, a little forced by circumstance, but now we have, let's call them completed lesson series for now, or parts of the programme, that we can pull out again in an instant. Also because our learners really appreciated them. But maybe we'll have to throw those out again sometime soon. In the sixth form there's a 'black literature' project, in the fifth form a project about 'love', and more along that line. Yes, I don't have all of that with me now, but I can show you. Yes, those projects really are the icing on the cake.*
> (1-2-8)

Mark claims his literary didactics have changed over the years, admittedly somewhat forced by the call for alternative didactic procedures that foster more independence in the learners. Nevertheless, Mark's interest in literature and the ease with which he assumes a central role in front of his grammar school classes was our reason to refer to him as the literary master. His belief in the potential of literary texts is first of all shown by what Mark considers his ultimate goal as a grammar school teacher of English.

> *The ultimate goal is to have learners leave this school with a fair amount of schooling in literature, next to having acquired quite a fair amount of knowledge and quite a fair knowledge of the English language as a means of communication, both oral and written.*
> (1-2-3)

Mark's ultimate goal reflects a core belief of his. The knowledge of and skills in English are at the service of realising an ultimate literary goal. To achieve the level of literary schooling he aims for, Mark was used to laying a solid foundation of grammar and vocabulary in the first three years of grammar school education.

> *It also meant that we, returning to that divide between first and second phase, have always said that we are a 'grammar school'. Especially in the first three years we're busy laying a very solid foundation. Language proficiency is important. Grammar is, was, I should say, at the centre of this. More so than it is nowadays. We followed this with a reasonable build-up of vocabulary and we were mainly busy laying the foundation for a sound house. That house would be handed over after the third form, unfurnished or vacant if you will. And the finishing touches, the furnishing, the extension, if necessary, those were taken care of by us in the second phase. And our taste was mainly literary.*
> (1-2-3)

Next to his experiences with literature at university, there appears to be a second formative experience he briefly mentions. Mark approaches grammar profoundly and methodically and refers to the importance of the first, three or four years of his career to manage to really grasp the structure of English and its rules of usage.

> *The first, three, four years, I was only teaching junior secondary classes, which gave me the opportunity to develop a firm grasp on English grammar and to learn to teach it effectively.*
> (1-2-4)

Mark adds that the concentration on grammar did not mean that there was not any attention to other skills he thought relevant. The course materials used by Mark increasingly focused on communicative skills.

> *And once again, I might be mentioning grammar too often, because I don't want to create the image that you, erm only had classes about grammar, you know. Depending on the course materials of course, and we used 'Learning English', I don't know if you can still remember that.*
>
> *I: Yes, I myself was raised with these course materials as a beginning teacher.*
>
> *M: Yes, they were succeeded by 'Unicom' and then 'Unicom plus'. Those course materials already encompassed speaking exercises, and listening skills of course. Yet, adding separate lessons to the course materials to test what are now called the basic secondary education (Bavo) attainment targets was out of the question at first, but has slowly crept in, because the materials have over the years been adapted to the BAVO targets. The sad result of this is that the grammar lessons will just continue in 'Unicom finals', the course materials we are going to use in the fourth forms, and that grammar will continue to be a focus in the fifth and sixth forms as well.*
> (1-2-4)

Mark states that the attention paid to the attainment of the Bavo-targets, of which he is not exactly a staunch supporter, prevents him from laying a grammatical foundation that is solid enough to concentrate on literature in his upper secondary forms. Mark justifies focusing on grammar and vocabulary in the first three forms as follows.

> *Grammar can often be annoying and, at a certain point, even deathly boring. That is why we prefer to focus on it in the first, second, and third forms, I think, because they're the age groups best suited for laying a grammatical, and also idiomatic, foundation.*
> (1-2-7)

The primarily structural knowledge and skills developed in junior secondary education used to be at the service of appreciating and evaluating literary texts in the forms to come.

> *And the skills, the language skills were at the service of literature. Which means that in the second phase we did everything in English. We spoke English, we wrote English, we listened to English, watched English films or videos, wrote essays, etcetera. Literature was the main course, it provided the texts, and that all with the thought in mind that it would offer you something valuable, something for the rest of your life, but then again, I am a romantic.* (1-2-3)

However, because grammar is on the curriculum in the new course materials, he feels his predominantly literary curriculum from the fourth forms onwards is endangered. I respond by remarking that most upper-secondary course materials include do-it-yourself grammars.

> *M: Yes, but also really specific exercises. And yes, those exercises are tested as well, of course. You can also see that in the tests that come with the book. And, once again, this is all a matter of just finding out as we go along, you know. We haven't got the experience yet., because we'll be using the revised materials for the first time. But what we do know is that we'll be working with fewer hours. Fewer even than what we've been used to so far. More hours will be cut next year. That means, specifically for us, that literature will be, well, almost decapitated from the programme, if you want to put it rather bleakly. But it also has to do with the profile that we've constructed over the years.*
> (1-2-4)

Mark generally assesses what his learners have learned from a short story, poem or novel by having them write literary essays. From the first form onwards, the learners are trained in developing the style and contents of the essays. The first and third of the tests that Mark selected for discussion in the course of the year were literary tests. This is how Mark gradually builds up towards improving the technique of essay writing and elaborating its contents. His approach is methodical.

> *And what I had actually forgotten was, that in the first, second, and third forms we actually already deal with learning how to write a book report. Very simple in the beginning, of course. We were very ambitious at first, by demanding one book report per semester. But then you drown in work. So we brought it back to one. And then we keep adding on to that. In what way do you list title, author, publisher, number of pages', then a little summary, and 'what was your opinion'. That's in the first form. In the second form we add your 'favourite character'. And in the third form we add something else, such as ten new words.*
> (1-2-16)

> *I: And you will keep doing this in the junior secondary forms?*

> *M: Yes, we will, once a year. At least that's our aim. It teaches them to, they don't get any marks for it anymore either, because it has to be satisfactory, so that they know how to do it. It's only been like that for a few years or so. We also do 'essay writing' at the beginning of the fifth form, just to show them how to write an essay. Then we study some poetry, at least I do, or a short story and then they're already doing all the work, all the thinking. Then they do the assignments in pairs, according to a certain handout, because we have to protect ourselves against spending too much time checking students' work. Then they write another essay again in class. The first essay isn't very interesting in its content of course, what's more important is that it should be organized in the correct manner, you know, uh //*

> *I: // The story's structure.*

> *M: Yes, 'state your intention', and all those 'arguments, line of argument, use of quotations and personal evaluation'. And afterwards they are given the task to write an essay based on a still unknown story or poem.*

> *I: And that's an individual test, I presume?*

> *M: Yes, or maybe they can do it in pairs again, because it is becoming more of a skill and it's also just to protect yourself because if you have nineteen classes, that really is a lot of work. In the sixth form we also still do that, but now we've chosen for a project work form, that they have to turn in with the*

second literature school examination, which is then discussed during the oral final exam. That works very well because then you don't have to evaluate everything all over again, because otherwise we always have a complete overview of what we have to check, including lay-out and certain items which total 100 points. Actually it's also quite nice for the learners, because they then know that their oral exams are based on their reading lists, which state what books they used for their projects. They know beforehand that those books will be dealt with. Well, that's quite convenient .

(1-2-17)

Mark sincerely believes in the added value of literature in the moral development of his adolescent learners. Learning about literature is learning about life. He wishes his learners to relate the findings of the literary assignments and the emotions they arouse to their own contexts and lives, and by doing so expand their horizons. Mark is more specific on the added value of grammar school education and the role of literature in the following segment.

I: One of the aspects you mentioned is that we all want everyone to read and that that really has an effect on people. That reading really makes a difference.

M: Maybe that's a good way to put it. An esteemed colleague of mine teaches the classics. If I remember correctly he answered a question from a fifth or sixth former. Who said: "Well, why do I have to do all this, you know?" Especially the classics are under pressure, if you're talking about 'use', I think. Then he said:
- Well what do you want from this education? Why do you attend a grammar school?
- Well I want that degree of course.
- Why?
Going at it like Plato, you know, that colleague of mine.
- Well, because then I'll be able to go to university.
- Why?
- Yes, well then I'll have the best chance of getting a good job later.
- Why do you want that?
- Yes, well because then I'll have the opportunity of becoming rich, making lots of money.
- And why do you long for that, all that wealth?
- Well, because then I'll have the means to travel a lot and see lots of things and buy everything and basically lead a very nice life.
And then he said
- And what will you do after you've travelled and have collected all the wealth and equipment in the world? What will you then do with your life?
Yes, and then he became a bit more quiet of course.
- Yes, well rest.
- Yes, and then what will you do?
And then the learner didn't have any more answers. And then he said. Well why don't you think about that some more. That's why you're attending grammar school.

I: The use of it all has come full-circle, with all the examples you just gave.

M: Yes, I quite liked it. I don't know if I told it exactly right, but that's beside the point. And, once again, maybe that has remained a romantic view. And, once again, there'll be plenty of learners that only benefit from English being offered as a tool for communicating, you know, that want to be trained as well as they can in speaking, listening and writing skills. But then again, I also think that through literature, you address certain issues that you have to contemplate anyway at a certain point in life, and that you might even want to

> *find the answers too, you know? And they aren't to be found in the final examination dossier, or whatever you might want to call it.*
> (1-2-12)

Mark's formative years as a university student were followed by his formative experiences as one of the two teachers of English at his present school. Mark is a self-made man. He was not inspired by collegial cooperation. The two teachers agreed on practicalities and on some final targets, but agreed to differ in other respects. As was the case with Joy, Mark first taught junior secondary education classes. It gave him the opportunity to develop his knowledge and skills in the fields of grammar and vocabulary. The course materials he used at the time helped him develop this knowledge and these skills (1-2-4).

Mark is ambivalent towards the course materials he uses. On the one hand he has learned to be critical of the contents and tests that go with the course materials, and on the other he benefits from the structure and ideas they offer the teacher and the learners in their weekly lessons. Mark, however, feels that the available course materials do not do justice to the grammar school learner. When I asked Mark whether he was to continue dealing with grammatical essentials, he said:

> *M: We'll keep on doing that. We've noticed that, and it makes a lot of sense, that course materials are not geared at the level of our grammar school learners. So you try to select the best course materials for a large comprehensive, for vwo-level. And then we sometimes add to that, under the pressure of school policy, if you will, which varies for each school, but for our school does have a large impact. For example the timetable for tests. This year, we're going to start with a period of four weeks of lessons, sometimes also three weeks, followed by a week. Of course we've found a lovely word for that, 'Quinduum', a period of five days in which the tests are taken, It is a kind of a test week, although lots of other activities can be planned for that week as well. Activity days or excursions, or, I should adds, time for meetings and mentor talks. In short, anything you can think of. And in fact because of that you're, well, kind of guided in a certain direction and when you know that you have four weeks time to finish a certain amount of material, which you always translate to a unit, a chapter. And then after three weeks, the system may require that you omit e.g. extra grammar you would have liked to offer, or a change in didactic approach, or you may even have to decide to omit a full chapter. On those occasions, you're actually, well, harming the grammar-school profile of which you've always been such an advocate. And when I'm in a bad mood, I do let them know how I feel about that.*
> (1-2-7)

Despite his criticisms of the quinduum and the text- and exercise books he uses, the course materials largely determine what the learners must learn and how they go about it.

> *We've always offered grammar as it has been presented in the course materials, without losing sight of the other skills. And that was quite visible in the tests, where if you had the subject of the continuous and the difference with the simple present tense, then you would ask them to translate some sentences, but you would also ask, for example, sentences they had to finish, and then make a sentence with, which we still do, with 'always', with 'often', with 'now'. So that on the one hand you would guide the learners, and on the other hand you would give them the opportunity to come up with something themselves. You know, like 'write a good sentence of at least eight words'. And then you'd give the first two words, 'next year'. Or 'if I had known this', … and finish the sentence, or similar exercises of that kind, besides testing their knowledge of vocabulary and testing whether they have or haven't understand the matters at hand.*
> (1-2-6)

The effects of the didactic approaches of course materials are not always seen as positive. The examples Mark offers are some overtly simple tasks and activities in the course books used in junior secondary education. concerning the materials used in upper secondary classes, Mark regrets that Dutch is increasingly used as the language of instruction.

> *And because in the upper forms English is the spoken language, everything happened in English, which, actually was also the case in the third and second forms… Even in the first, the tasks all went in English. But now they're already deviating a bit from that, right, because they also do that in those course materials, you know. Yes, so we're taking a step back.*
> (1-2-14)

Mark feels that this loss of English is an instance of putting the cart before the horse, which he finds hard to put up with.

So, Mark developed as a teacher by an interplay of input from the course materials and his own beliefs in laying a firm grammatical foundation in the first three forms, which is subsequently applied to the interpretation and appreciation of literary texts in the upper forms.

Mark never mentioned he was inspired by the views and ideas of his colleagues. When he started teaching Mark only had one colleague. The two teachers of English saw eye to eye on essentials and some final targets, but agreed to differ in other respects. Mark believes in teacher autonomy and regrets his paradise lost in view of the changes he does not see as improvements.

> *And once again, keep that in mind very clearly. Of course I come from Paradise. I've had the privilege here to always work with one single colleague. Because there was no cooperation. Not because there didn't have to be any, but well, we both sold the same product and we just kept an eye on the basics. And that meant you knew what the final goals were and you almost always had the same learners from the first form through to the sixth.*
> *Or from the first till the third. And I started with three first forms with 21 learners. And now, well, now there's been a scaling-up of , I can't even describe it. The renewed second phase, I think, is going to demand an awful lot of effort and flexibility from both teachers and especially learners to have the school run fairly smoothly. Because, in fact, we are to start tomorrow, but we're in classrooms where nothing has been set up yet. Because they're still painting and they're still whitewashing and I don't know what else and there isn't any equipment yet at all.*
> (2-1-10)

At this particular stage, we interrupt Mark's voice for a short while. Among the notes that Mark took while reading the interview-guide were two transitions he wished to mention in the course of the interview. These changes are indicative of the clash he feels between the demands of the innovations and what he profoundly believes in as an educator.

We have already seen that Mark far from welcomes the innovations of the second phase. He objects to its rigidity and large-scale approach that puts great demands on the staff, learners and resources of the school. Mark is equally critical of basic secondary education. He sees more drawbacks than opportunities of the curricular and didactic changes he is expected to implement and support. Mark considers his own beliefs and didactic approaches as viable alternatives to the sea of change. Mark feels his beliefs are endangered by the attainment targets of basic secondary education and the renewed second phase. The two curricular and didactic innovations have led to transitions he seriously objects to. In the first interview, he frequently refers to these transitions. One of these changes is the aforementioned spread of essential grammar and vocabulary over the full six years of grammar

school education. The other transition pertains to the changed role of the teacher. Mark feels both of the changes have gradually been incorporated in the course materials he has used over the years. In the paragraphs below, we will provide more detail on Mark's objections and criticisms.

In forms 1-3 , Mark was used to working at a knowledge-based foundation of English grammar and build up and gradually extend the learner's mastery of vocabulary. He had always stimulated his learners to go for perfection in these areas. However, the attainment targets of Basic Secondary Education primarily concentrate on skills rather than on knowledge and clearly focus on learning by doing.

He mentions two dangers of such an approach. On the one hand the learners have to do communicative tasks that are way too simple for his learners, such as ordering a loaf of bread at the baker's. On the other hand the learners are asked to do tasks for which they simply lack the necessary knowledge of and skills in English, which results in compensatory behaviour and a kind of English he cannot support. Mark strongly feels that both the tasks that are too simple and the tasks for which the learners lack the required level, take away valuable time from building a sound grammatical and idiomatic foundation in junior secondary education, which should, primarily, be knowledge-based. On top of all this, there is no longer time for all of the extras he and his colleagues used to offer the pupils of forms 1-3 in the fields of grammar and idioms.

The result is that basic grammar and idioms increasingly have to be part of the upper secondary curriculum, because without this foundation the learners cannot communicate adequately at the level he expects them to. The level Mark is after incorporates the knowledge of and skills in English needed to study, discuss, and report on literature and literary texts.

The second transition Mark refers to is the shift in the role of the teacher. He regrets the loss of the role of the teacher as an ex-cathedra educator, who sets examples for the learners to copy and is very much in control of what is learned and how this is done. Mark uses the Dutch words "*onderwijzer*" en "*leraar*" to refer to the original role he used to have as a teacher. He feels that in the second phase his role is sadly reduced to that of an instructor supervising learning achievement from a distance. He uses the Dutch word "*docent*" to refer to the new role that he feels is now forced on him.

Some twenty years ago he was an educator able to flexibly train and test his pupils in English, and prepare them for life and further education by focusing on literature. He stresses the exemplary role a teacher has. The teacher introduces a literary subject in English and provides two or three examples of how the subject can be approached. The learners take notes, and afterwards take over the role the teacher initially had.

Now, Mark feels his role as an educator is marginalized to being a mere instructor or facilitator, who is almost forced not to interfere with his pupils' work. He fears that these developments will affect the high levels of English and the development of academic minds he has so far been able to achieve.

Mark mentions the influence of the course materials on both of the transitions he has observed taking place in the junior and upper secondary curriculum. The materials have changed over the years and have increasingly focused on preparing the learners for achieving the junior and upper secondary attainment targets and passing the national examinations of foreign language reading skills all secondary school learners take in their final year. The new range of course materials has been designed to stimulate learners to partly work and learn on their own.

In the Netherlands, teachers can select from a wide variety of course books. Mark and his colleagues had opted for *Unicom Plus* (forms 1-3) and *Unicom Finals* (forms 4-6), course materials known for a relatively firm focus on grammar. The revised version of Unicom Finals had just been published. It is going to be used in the fourth forms. Mark says that he has always offered grammar by way of the

course materials. This means that the gradation of grammar, the didactic procedures suggested and the tasks the learners are expected to do determine how his lessons are structured and the curriculum is built up to a great extent.

Marks sees the two transitions as deteriorations. He is afraid that the typical grammar school graduate will be lost by the two main changes taking place. He fears both his school and his pupils will no longer be distinguished from large comprehensives and their learners in the ways his school and his learners used to. When he expressed his dislike of the compulsory national tests to assess basic secondary education targets for English, Mark metaphorically describes that he would like his learners to reach the most southern part of one of the more beautiful provinces of the Netherlands.

> *I never believed in BAVO tests. And now I'm talking as a grammar-school teacher. Maybe seven, eight years ago, when they held those meetings, I already mentioned in good company that well, what we want is, maybe I even told you this once, you know the train, you know secondary education, leaves from Nijmegen, and some schools go to Cuijk, some schools go to Venlo, some go to Roermond and some, including us, travel as far as Maastricht. And the BAVO tests, and then I'm not even talking about their contents, require all of the learners from vbo- to grammar-school level, to get off in Cuijk and show their valid Ids, and then the learners who have to travel further must hurry back on the train again. And that really is a horrible waste of time for us. Because we want to get to beautiful southern Limburg as quickly as possible. But well, the attainment targets of basic secondary education do not allow this and are in the way of schools that really rise above the rest, as our school does, with its what can be called grammar-school profile.*
> (1-2-4)

Mark feels that the compulsory tests that Dutch secondary school learners have to take at the end of basic secondary education are an absolute waste of time. He justifies his view with an example. Some time ago, he had discussed his abhorrence with the tests with the School Inspector responsible for his school. The test items were so deceptively simple for his grammar school learners, that they tended to make mistakes.

> *M:…and yes, I remember clearly that I told the inspector at the time: "Really, it actually says "What does this sign mean?". And then you'd see a person, you know, doing some activity with a line running through it, and a caption that said "No smoking" , "No swimming" or "No cycling", just to mention a few. A grammar-school learner would almost think 'It can't be correct to go for no swimming; it's got to be a trick question'. 'Yes', the inspector said, you're completely right, but you must test it. And I replied: 'Yes, but I'll be damned if I do'. I said: 'That's all fine and dandy', because I'm not mad, I wasn't hired to do this. Look, I'm not at a, you know, I'm at a grammar school, see, and so this is my objection. And I still fully support it. The BAVO is dying out, especially for us, and paper is patient.*
> (1-2-9)

Mark's words would prove to be prophetic. In the revised plans for basic secondary education, more autonomy will be given to the schools and teachers regarding the curriculum and didactic approaches in basic secondary education. Mark is equally critical of the second phase. What exactly did Mark prophesy about the second phase?

> *M: If you want to speak evil, then you'd ask how much this actually costs? How much does this really cost ? All those people there, yes, the real practitioners I always call them, they must be consulted. But no, there's no time for that. Now you have to do this, because practitioners didn't actually have enough say in the matter, I believe, and maybe even get blamed with phrases such*

as 'But you had the chance to have your say through all kinds of different channels.' But the daily practice, well you just saw it before we started, it's so demanding, and at a certain point you've just had it.

I: The days when you used to arrive home at around two, (M: No) now it's always late (M: Yes) in the afternoon. Yes, that's pretty clear. So, the workload has definitely increased.

M: How can you keep it up, you know, because that's the question of course, also from the learner's point of view. It's all becoming so serious. It has to be serious of course, but every hour as a learner, you have to convey the message to your learners, utilize it. You have to go to the library or to the study hall, because otherwise you have to do all these assignments at home.

I: Words that you've emphasized again: "use, planning, efficiency, results//

M: // Yes, they're still teenagers and groups of learners who you can't treat the same as students. They also experience periods when they're not feeling well, or whatever. That's the big mistake, I think. Apart from, well, some teachers sometimes. I mean it has repeatedly happened that I thought it was quite appropriate for learners to say to me 'Yes, but sir, now we're talking about..'. Sometimes it was important, sometimes it wasn't of course. 'Now you're always talking or we're doing this' a nice conversation perhaps. 'But you always say that what you can do in class, you don't have to do at home.' In other words, just shut up, sir. Let's begin. And of course they were totally right. But are there opportunities these days to follow up on current events? Is there room to have a bit of a laugh when it's some girl's birthday, or about NEC [soccer club, ed.], or I don't know what? And that's the icing on the cake, or what do you call it, I think. But maybe I'm a bit old-fashioned. I love that and at the same time I was basically the planner who kept his eye on where we had to get and you were still allowed, didactically speaking, to say, well, class 2a - when you could still plan your own tests- Well, class 2a hasn't really got that yet. I think we will need an extra lesson for that. And 2b, well they're pretty clever, they can go on. That's no longer possible.

I: There won't be any room for that this year//

M: // Didactically speaking, you are just adding a lot of water to the wine. You have a schedule, where this testing moment is assigned to you and then it might be that your test is on chapter one, but you're already halfway through chapter two or, because of the cancellation of lessons or something, you haven't really had enough time in a certain class, to really teach them well, but you still had to it, with all the necessary consequences. So, didactics is no longer a word in our vocabulary. We're all becoming an educational organisation//

I: //In which you function as an organiser, and no longer as an educator, in the sense you pointed out at the beginning of our interview.

M: In that sense, yes, yes. I do think so.
(1-2-10)

Mark feels that the voices of practising teachers have been virtually absent in preparing for the second phase reform, basically, because teachers lack time. He experiences a considerable increase in his workload and objects to the impersonal and business-like organisation of the second phase, which turns schools into organisations that has its learners primarily work on their own schools. An important drawback for him is the lack of flexibility and opportunities for differentiation. Mark foretold a bleak start.

In discussing one of the main goals of the second phase innovations, i.e. ensuring that the secondary school graduates are more successful learners in higher education, Mark comments that education based on learner autonomy can be highly selective.

> *I: So there's no selection in the second phase?*
>
> *M: Well, I don't know, but it is a consequence of the policy that we've enacted. Once again, keep this in mind and write it down correctly, because it should be so that in this way, the second phase is in reality also a cut-down, which should prevent us from delivering learners, vwo -learners, who will fail their academic education. That's the idea behind it. That there should be fewer people occupying positions, space at a university. Space that is lost, because they can't plan or they can't handle the responsibility.*
> *Whereas we have, because of our system, and of course also a little bit because of the age group, but we have always kind of held the learners' by the hands. Although, well, we only hold their pinkies now and most of them we don't even have to hold at all. But that 10 percent, 20 percent that make it now, because we did push them, they're not going to make it at the university. And so now we're also saying that, well, the new system demands more of the learner regarding their own responsibility and if they don't show they have it, we simply drop them. Then they should just go to the havo-level. Then they can fail there. Is that what we are doing? That's the question I've put before the school management in meetings. Or do we actually still create a safety net? … Only time will tell.*
> (1-2-30)

Next to the two transitions Mark has highlighted, he mentions two additional drawbacks: the marginalisation of literature in the curriculum and the reduction of contact hours with his learners. Mark deplores the marginal role of literature in the second phase. He frequently states that upper secondary education without literature is a decapitated form of education. He attempts to hold on to a literary syllabus as long as he can. Another drawback he points at is the loss of contact hours with his pupils in forms 4-6. A second phase teacher will see his learners less often, because the learners are supposed to work on their own.

In the first half of the interview, Mark also expressed some core beliefs on language tests and foreign language assessment. He puts forward that he is critical of the tests that come with the course materials.

> *They are of a level that does not allow for discrimination and they are too easy. On top of that, they are too extensive. Four, five pages that you all have to copy.*
> (1-2-13)

Whether Marks' criticism results in adaptations or even the construction of tests of his own, will be taken into consideration when we discuss the tests he is going to select for discussion.

Mark is interested in some of the microskills that underlie oral and aural skills in English. He claims he has always assessed pronunciation and welcomes sound discrimination exercises.

> *What I have always done from the beginning is to practise pronunciation and to record some sort of system for that. A plus or a minus or a plus minus. Those that got a minus would be heard more often than those with the plus. Well that's still going on to this day. There's a larger variety of exercises, but that actually has to do with the course materials that are on offer. The core remains vocabulary and grammar in the lower forms. But, a lot of additional exercises are suggested by the course materials. I can remember a lovely exercise in Unicom, not Unicom*

> *plus, about which words are not stressed on the second syllable. Those are wonderful exercises.*
> (1-2-13)

Mark believes in error correction, when classes do badly on a particular test. At the start of his career, errors were corrected by Mark himself. Later on, he had his learners analyse the errors and mistakes they had made. Now, he only occasionally has his learners do error analyses.

> *In the very beginning, when I still had eons of time, because I started with 14 hours, I was so mad, I now say, as to keep a little note book of each learner, or to jot down in a note book, what grammatical errors they would make. Can you imagine? I also corrected each mistake, which isn't psychologically correct, it seems. Meaning to let the learners know what the mistake was, so that we could get back to that. That was really back in the days when you were still king of the castle and you would agree with the class when you were going to test, at some point in time when you were all ready for it and then, for instance, if 1b didn't quite master the grammar yet, then you knew, well, next year, in 2b, we'll just continue where we left off and well we just can't do that any longer.*
> (1-2-14)

Interestingly, Mark used to add process-oriented questions to tests he administered, by having his class answer questions such as 'How did you prepare for the test?', 'What did you make of the test?', 'What mark do you expect to get?'.

> *And what I actually did from the very beginning was to ask one simple question: 'Any remarks you want to make?'. Where I often asked them in the first and second forms to write down what they thought of the test, how they had studied for it, what mark they thought they would receive. Rather a moment of reflection.*
> (2-1-13)

In view of assessing a learner's communicative skills, Mark feels that learners should be tested at the end of the learning process: "I really believe that you should assess a learner at the end" (1-2-9). Thus, he seems to consider language assessment primarily as a summative process.

In chapters 5 and 6 we already explained that the score on the national reading comprehension test determines half of a learner's final mark for English. Mark feels that the mark a learner gets for his national reading comprehension test at the end of form 6 is not a valid indicator of the level of English the candidate has mastered. He is of opinion that his learners do well on reading comprehension, partly because of the attention that has been paid to the interpretation and understanding of literary texts.

> *But as I have always said, I have never been very interested in scores on 'reading comprehension', because once again, to put it perhaps a little too sharply, it doesn't tell me anything about their English. The real mark for English is the school examination mark. And it's just that reading comprehension is tested by way of English texts. They could also have been in French, or German. It's a certain skill that you test. Besides, you're also testing their concentration levels, and their general knowledge, because people who already know about the Gulf War or Greenpeace, well, you name it, are on the right track already. And people who don't know about those things yet, are at a disadvantage and that really doesn't say anything about their levels of English. In a way it is unfortunate that grammar-school learners often achieve scores on the national examinations you can take your hat off for. Scores of '9 point so-and-so' are not exceptional, even though scores tend to get a little less high of late. I think these high scores are partly due to the attention paid to language training by way of literature. The scores on the national examination tests make up*

fifty per cent of the final scores for English. So people start to think you can do with less time with results like this. And then you think, well, guys, where is the appreciation for what you do, you know.
(1-2-12)

Arguments such as the ones presented by Mark, in which he challenges the validity of the national CITO reading comprehension tests and glorifies the validity of the school examinations, tend to defy one to start an interesting and possibly somewhat heated debate. However, in line with the objectives of this study, there were no debates with the respondent teachers in the year of data collection. Such debates are postponed a little for now. In chapters 11 and 12 we will discuss how the teachers' core beliefs on autonomy, communicative proficiency and language assessment are reflected in their assessment and evaluation practice in settings where the learners are expected to learn to communicate in English.

There appear to have been two main formative experiences for Mark. The first and more often mentioned of these are his experiences as a student of English at university. It made him appreciate the materials, didactic approach and assessment practice of the subject of English and American literature. When he became a grammar-school teacher, his academic experiences appeared to have influenced his choice of literary texts, literary didactics and assessment practice. Mark only briefly, yet significantly, refers to a second formative experience. Mark feels that the first three or four years of his career that he spent on teaching junior secondary classes, gave him the opportunity to develop a thorough understanding of English grammar and learn how to teach grammar effectively (1-2-4).

In the next three sections we will report on Mark's definitions and interpretations of learner autonomy, communicative language education and language assessment and evaluation.

7.4.2 Learner Autonomy

Mark distinguishes three components in his definition of learning on your own.

- *planning skills, and the discipline to put the plans into practice;*
- *these planning skills are related to:*
  - *the learner's awareness of what to spend more time on, the skills they can rely on, and what they already know;*
  - *the learner's ability to balance learning activities with doing tasks and assignments;*

M: Yes, those are notions that require redefinition time and time again. Because learning on your own, to me, means that learners are taught to plan. It relates to what I said about the programme in the fifth and sixth forms, when learners start to understand what they have to spend more time on, what skills they can rely on, and what they already master. Teaching the skills of planning and in such a way that the learners can actually realise what they planned to undertake. Whether they do so isn't something you can really check. You can push them, though. The skill of, well, everyday things such as, which I never did well either, balancing learning and doing. Learners are prone to do exercises one after the other. Then you can put a full stop at the end and you're done. You know, that's an ongoing battle and it doesn't matter how you put it and what theories you have about it, but it's always the same.

I: Learners are product-oriented, in principle.

M: Yes, but not just learners, I think we actually all are.
(1-2-27)

Mark feels the primary component to have learners regulate their own learning is to teach them how to plan, and encourage them to stick to their planning. He sees a clear role for himself as a teacher here. However, he adds that successful planning is also related to the learners' awareness of what they already know or are able to do, which helps learners decide what still needs a little work. Mark feels such awareness is developing in the fifth and sixth forms. Thus, what Mark implicitly says is that self-assessment and self-evaluation are qualities that need to be developed in order to effectively plan what is to be learned or done, at what particular time. Because he feels that human beings are basically product-oriented, he is convinced that fostering learner autonomy in the learners is a job, and having the learners develop awareness of the learning process 'an ongoing battle'.

Mark sees learning with learner responsibility as carrying out what learners have been taught. The actions and activities become more and more learner-driven. Learners can be taught to hold on and persevere. He feels that mentoring is essential in this process. There are many ways in which you can coach learners. Mark claims his style is not to be on top of everything the learners do all the time. However, if a learner fails to take responsibility, Mark certainly exerts his authority.

> *Learning with learner responsibility then means that, in a certain way of course, you have the responsibility of doing something that you actually have been taught. A skill, which you actually carry out. That is more and more up to the learner himself, although of course there is always still the mentor who, such as I did, had talks with the learners the past year, because of course you get more and more of those. And you have to// Everyone does it as he sees fit, but I am not someone who is right on top of things. That's not what I like. But in some cases, yes, you have to check on them each and every lesson. I think that's quite contrary to what we're doing, but I can remember last year when I had a third form and the learners were supposed to work ahead on one unit from Unicom Plus, working towards the date of their test. One or two lessons before the test, I suddenly noticed that the learner sitting in front of me still had to do quite a few exercises. I went up to him, thumbed back a few pages. He hadn't done a single thing. So I have the rule that after we do some things together, like reading some texts or explaining something or doing a language exercise. Say, we have 25 minutes left, then I'll say, well guys, now it's quiet until 12 o' clock and then you can deliberate. And when I mean quiet, I really mean quiet. It's my responsibility to guarantee that. Old-school style. Noise and learning are irreconcilable. And then I disrupt the relative quiet by saying 'And what's this then?'. (louder) 'What's this?' Putting on a show, and maybe acting it, but not in the perception of that particular learner//*
>
> *I: //If it could have been quieter, it would have been.*
>
> *M: 'What have you been doing? And then in two weeks, well. And that really works, because all you have to say is 'I think that for my next lesson I'm going to check whether there are any more slackers like this in this form. Because this just makes me feel sick. Have you totally…'. Of course, you can hardly conceal your laughter on the inside, but in any case. Right, then they're all at it again. Apparently, you do need that.*
>
> (1-2-28)

Mark feels that the learner is responsible for transferring what (s)he has learned to tasks and activities with dedication. This requires some discipline and perseverance. If these qualities are lacking, Mark has no problems coming down on the alleged culprit like a ton of bricks. What Mark stresses is that teachers have to make sure that learners do not abuse the autonomy that has been given to them.

7.4.3 Communicative Language Education

When Mark is asked to define communicative language education, he first asks me whether I want him to digress on the role of the teacher in communicative language education. Because it is the respondent's interpretation we are after, I encourage him to pursue that line. He sees himself as an instructor and a provider of stimuli who cooperates with his learners and subsequently, has his learners cooperate in pairs or small groups. Mark then touches upon philosophical and cultural dimensions of communication by stating that life is all about communication.

M: Erm, communicative language education. Erm. You mean what the teacher's role is in all of this, or?

I: You could see it that way, or what's expected from the learners//

M: // A teacher is required in education. I think that the teacher's function in all of this is that of an instructor and admonisher. A certain level should be taught, accomplished and practised, by way of example, but also through instruction and practice in cooperation with the learner or learners, and, as the next step, through cooperation between learners, dependent on the form they are in, and according to the principle that life, and now I'll become a little philosophical, is all about communication. In whatever fashion. And then we're focusing on language. 'Let's communicate or die'; John F. Kennedy in his inaugural speech. A nice title for the resulting paper. And that can be done both orally and in a written form. I already explained that earlier. (1-2-25)

It was a definition I had not anticipated. I tried to put Mark on the track of English use and usage by summarising some of the views he had expressed so far. In his response he linked communicative language education explicitly to grammar teaching and language awareness.

I: Yes, well I heard that there are two very clear aspects for you. When you talk about communicating, on the one hand you're talking about teaching, about conveying a message, and also motivating your learners. On the other hand, communication requires the use of English. We talked about that as well.

M: Depending on their form.

I: Yes, all right, depending on their form. And at a certain point, there is transfer, because the learners start cooperating and working independently.

M: Instruction in the sense that, but then you're talking didactics, not as in let me show them and tell them about it. No, but in cooperation with the learners, with the audience, you're trying to clarify the rule or the difficulty. And it's not that I'm like the verbal overview of a certain chapter about grammar that you can read at the back of the book. Together with the learner, you're trying to clarify things, often in a contrastive manner, especially in the junior forms, you know. And .. I think that's something that should also be stressed in communicative language education. I'm not doing it on my own. Neither am I merely showing them how, but I am the one to initiate and I definitely am the one who thinks of and presents clear examples. I may write on the blackboard, say, 'I've lived here for five years' and ' I lived here for years'. You know, there is a difference. (1-2-26)

I: You ask learners to think about the differences between the two.

M: Another example is 'This is John's picture ; This is a picture from John; This is a picture of John'. It still occurs in a number of course materials, but I do think it has to do with, hey, what are you actually saying? You know, and I'm

not giving it away, but, hey, what do you think? Then in the end we will write that down in our notebooks.

I: So you provide examples, specific, clear examples of language usage, don't you. Of the English you can hear around you or that//

M: Yes, yes. And of course you use the tapes and …..
(1-2-26)

Mark does not really offer a working definition of communicative language education. In his interpretation, he first thinks of his role as a teacher, who via direct instruction controls the learning process and attempts to have the learners understand difference in usage between English and Dutch.

We will now focus on Mark's views on language assessment and evaluation.

7.4.4 Foreign Language Assessment and Evaluation

As we have seen when we discussed Mark's formative experiences, he had already presented some of his core beliefs on language assessment. Towards the end of the interview, we more particularly focussed on assessment and evaluation. We will learn more about or recapitulate the knowledge, skills and understandings Mark considers essential for his learners to learn, his definition of an effective English written test, and his interpretation of test washback.

Essential knowledge, skills and understandings

Mark had already indicated that the essential knowledge and skills were determined by the contents and gradation of the course materials he uses (1-2-6). Neither does Mark seem to focus on one language skill in particular. The four language skills are generally offered as they have been presented in the course materials. All of the grammar, vocabulary and language skills that the learners study and practise with are meant to lay the solid foundation needed to read, understand and discuss literary texts, orally and in writing. To this end, the gradation of the texts, grammar and skills offered in the course materials is generally followed meticulously.

With the exception of spelling and careful formulation, Mark has not singled out any particular grammatical knowledge in the first interview. The examples that he offered were related to pronunciation, verbal tense and form, and having the learners finish if-clauses in appropriate ways. Below is a segment that illustrates the importance he attaches to appropriate spelling.Mark mentions precision in spelling and formulation as two skills he would like his learners to master in the course of their secondary school years. As a lover of literature, Mark detests bad style.

M: …spelling is very important to me too. Precision, things that seem to become less important in day to day life.

I: With a focus on all of the language skills, spelling is not really or hardly//

M: // yes, or precision. it's really sort of a reflection of our society I think. Like, well, you know what I mean.
Take e-mail for example and you'll see what worthless style people use in their sentences.

I: That's an excellent example, yes.

M: I really hate that, but then again I'm a romantic who appreciates the way you can state things, especially on paper. You have to consider that you have to pay attention to the way things are being said, because you are documenting things. That goes squarely against today's fleeting society where people tend

to delete. Delete, that's actually a button, with which they can delete even more. I think that's a shame, but then again …

I: We'll talk more about effects in a later stage. If you find spelling important, and if spelling plays a part in your assessment , you should at least see a change with some of your learners of course. Whatever it //

M: // but then again you're not really autonomous. Because you have to find a framework, within your department, which isn't very difficult, but also within the school really. A framework that is supported by your colleagues, but sometimes if you see how teachers formulate certain things I think, well, yes, maybe it is a waste of time and energy to pay attention to spelling and language.

(1-2-23)

Mark was also asked to interpret the role of insight when learners learn how to communicate in English. Here is how he responded.

M: Yes insight is a term that you can't just explain one-two-three, unless it has been put into a particular framework, you know. Literary insight is different from insight into how a language works. And you could continue along that line. If you're talking about literature, literary knowledge, then I prefer to test it in an oral form. And I really find the personal aspect very important, in the sense of what do you think about that and why is that what you think, can you substantiate that. How do you see that character or that character in the novel, what does he have to say to you or what emotions did the author manage to arouse with his story, for example.
Grammatical insight as such is nonsense. Although, on second thoughts we may be concerned with it a little. It's kind of part and parcel of grammar school education. An example would be the connection between adjective and noun, and verb. So you are creating it a bit, but when you look at parsing, which can be important at times, then in a way you assess that by way of small tests. And then there, then it's mainly a question of knowledge. Yes, in a way. Although I sometimes think that it's better not to literally take sentences from their notes, but to force them to think a bit more. For example 'Can you finish the sentence: If I…' Then you know that they should use a main clause, then maybe they could use 'will'. Or you start with 'Next year..' and finish the sentence. Then they have to think about it. I don't know if that's what you call insight, you know. That signifies the future tense, so then they have to use that.
(1-2-24)

Mark relates insight and understandings to the type of knowledge or skill they relate to. Literary insight differs from understanding how a language works. The outcome of literary insight seems to be that a particular interplay of knowledge and skills helps the learner express and justify an opinion of his/her own.

Mark first has doubts about the existence of grammatical insight, but on second thoughts says that it might play a role of some kind, e.g. in connecting particular semantic relationships between adjectives, nouns and verbs, or in parsing. This type of language awareness is created by 'forcing' the learners to do a bit more thinking.

An effective written English language test

This is how Mark defined an effective written language test of English in upper-secondary classes.

M: An effective written language test is a test// Yes, then I need to distinguish between junior and upper secondary forms, I'm afraid.

I: Shall we focus our attention on the upper forms in connection with your fourth form?

M: Yes, upper-form tests relate to the level the learners have attained. It concerns tests in which a learner has the opportunity to, next to, say, what's been learned, show his or her creativity and his feeling for the language and his ideas in English as well as they possibly can. And it's a written test we're talking about, aren't we? A written test, in which you can state in concrete terms// Well, this is what I'm going to do in the fifth form tomorrow, because they're new learners for me. Write a letter of introduction, in which I ask them to introduce themselves to me, and to write about their history regarding English at this school, what their strong points are in English, and their weaker ones, and what they think of the programme that I have proposed to them. And then they can also write about their holidays. I will tell them to 'write the letter as if you are talking to me', and ask for a minimum number of words, because usually I have to limit the length. Otherwise, checking the letters becomes an awful lot of work, and the letter of introduction is really a way of becoming acquainted with the learners. In more than one way. But I think that should always be included. And I also often tell my learners, it even used to be included in idiom, 'Diligent and hard work will always get you passing marks for English'. Because the more you write, the more mistakes you make. But then you have to find the balance, you know, because in what way can you just open the gates wide and tell them to write what they like if you adhere to the principle that the result should be of a reasonable quality. That's important of course. (1-2-22)

According to Mark, effective written tests first contain what the learners have had to study. In addition, the test have tasks or assignments in which they have to creatively use the texts, grammar rules or idioms that have been studied, practised and learned
(1-2-6 ; 1-2-22; 1-2-23).

Interpretations of washback

Mark cannot imagine a school without tests. Tests provide the incentives people, and not only pupils, need in order to get some work done and achieve results. Tests enable both the teacher and the learner to look back on what has been or should have been learned.

I: Well, Mark, we have now arrived at the final two question categories of this interview. A very important one is the significance and potential influence of test items and tests to the learners. What is the role of tests when learners learn to communicate in English? … What do you think?

M: Phew, now that's a difficult question because I think, well, erm, I don't know if you've read that Robert Pirsig book, Zen and the art of motorcycle maintenance?

I: Yes, yes, I have. I'm a motorcyclist myself, so it was almost required reading for me.

M: In any case, not that he deals with it specifically, but, could you imagine a school without tests? Can you? Wonderful questions to philosophise on, also with learners. I don't think it's possible. I think tests are essential incentives to human beings, .. not just to learners, but to human beings in general. If you know there's going to be a moment of reckoning, you actually do the work you are expected to do. And even if you find tests and testing meaningful, if you know there won't be a test with consequences, then, apart from the ones who are really inspired, motivated, or whatever, you will not perform as well as with tests that will be scored and graded.
So it can be very stimulating, an affirmation of what you've learnt. But it can also be the other way around, of course. That people totally miss the target. That they'll write down that they have studied for a nine and expect that's the

mark they're going to get. It sometimes happens in the first form and then they end up with a 2. It sometimes has to do with their plans of approach and at times also with overestimation of their capacities and abilities. So, yes, tests do influence how learners learn. They are essential in that respect. Tests are also a means for the teacher to find out how he's done. That used to be more so the case than now, you know, an indication of how well the learners have been doing. This goes for the learners as well. They're also getting all kinds of reflection tests and the like in their course materials these days, and they are asked to keep personal logs and diaries. How did I do? (1-2-31).

Mark believes that tests may positively influence the ways in which learners learn, and the ways in which teachers teach. The incentives and feedback that tests provide are indispensable, in education as well as elsewhere. Marks adds that learner progress is increasingly assessed by way of reflective, or diagnostic tests and by means of personal logs.

In the next segment, Mark comments on marks. Nothing succeeds like success.

M: Yes, yes. The results of tests are important. Results, the marks they get. I think that's very// It's really a vicious circle. If you do a test and you get a nice mark, you'll be more interested and motivated to do well again next time. And the teacher becomes so much nicer and the subject becomes so more interesting and so on and so on. Without tests, you're in another vicious circle, but surely one that is a little more negative.

I: Yes, so there's a relation between motivation and test results?

M: I think there is. It's nonsense to say that learners choose a subject just because they like the teacher. No, it probably has something to do with it, but as the teacher becomes more of a master, then that should have less of an effect.

(1-2-32)

Mark directly links positive results on tests to learner and learning motivation. Mutatis mutandis, absence of language tests will lead to underachievement.

At the end of the interview, Mark went into plans for the school year to come.

7.4.5 Mark's plans and intentions for the school year

Mark will attempt to maintain what can be maintained from the old syllabus in the new syllabus he feels forced to develop. He plans to balance the old and new syllabi. He realises a lot of work will have to be done. It will be hard to offer any extras on top of what the course materials already offer the learners. The new quinduum set-up forces Mark to plan and arrange the tests, and so does the PTA he and his colleagues had to produce.

Mark is confident that talented learners will survive in a set-up in which learners largely work on their own, but is less sure of the average learner who is more in need of teacher support and control. Will the learners be able to cope. Will the teachers? Only time can tell.

I: Do you have any special decisions and resolutions for that fourth form for this coming year?

M: Yes, I wrote them down, but the future will have to tell how much we will be able to realise. To find the right balance, that's what we're looking for. See what the learners can handle in the time we've got. We want to see what we can retain from our old programme and then it will just be a question of comparing the two. And because I'm now a mentor in the fourth form, I will

have a better understanding than before, when I was only involved with classes 4a, 4b, and 4c as a teacher. Insight into the weight and workload of the programme may cause the English teacher, with all his wishes, demands perhaps, to add water to the wine. I'll be keeping in mind, though, beware people, we're starting to slip here. It might not be a case of decapitation, but I think it will be a step towards it.

I: So in that sense it's going to be another very important year. Especially since it's the first year of the second phase.

M: Yes, it's really crucial, yes, yes. We'll have to be more in tune with each other, you know. Within the department yes, but also within the whole array of subjects they have. So those are really our resolutions for now. And yes we don't// There's still room for variation, as far as testing goes. At least we put it into writing as such. We will have tests on every two chapters from Unicom Finals. And then, if we can, we wish to include a test about a short story, in class. And a test related to a poetry course, to give it a name, which assesses the basic notions according to some new poem and then we have that paper about a novel that they will read in groups at the end of the year.

I: So then you're talking about the final goals for this year?

M: Yes, yes, because we have all had to write them down.

I: In the PTA?

M: //In the PTA, but yes, and then we end the year with a dossier test on reading comprehension. Yes, I wasn't in favour of it, but then apparently we have to, and we happen to have chosen for reading comprehension as the skill that is focused on. The test has a weight of factor one, so it won't be very important for the final mark.
Actually, hmm, not very logical because you're dealing with a language and I think I started that way, with the road that is to lead to Maastricht. And then you're going to stop in Venlo to check how they're doing with reading comprehension? And there's a definite chance that people who might just get off fresh as a daisy in Maastricht, will huff and puff and score bad results here, whereas they could really improve in the two years still ahead of them. Learner with bad results may feel weighted down, you know. Despite its factor-one weight, it may still have a psychological effect. But there's really no alternative to reading comprehension. With the standards with which we assess writing skills, we cannot possibly assess their writing as a school examination this early. It's the same for listening. Even though listening has been practised before, a test at the end of the year would really have to be made easy, but that's rubbish. So for language it's// … No what we had was fine. Plan assessment as late as possible, and bring back the weight of reading comprehension.
(1-2-33)

This is where the first interview with Mark ended. In the school year to come, I was to monitor Mark's assessment and evaluation practice in one of his 4 *gymnasium* forms. In the last section on Mark, we will provide some information on the tests he was to select. In chapter 9, we will report on these tests in more detail.

7.4.6 Mark's language tests

Mark selected three tests, which we discussed elaborately in the course of the school year. In correspondence with his core beliefs, Mark chose two literary tests and a test on two chapters from *Unicom finals,* the course materials he used.

The first test was a literature test on a short story the class had not read before. It concerned *A Day's Wait,* written by Ernest Hemingway. Judging by the copy that was

given to me, a copy made of a good-old black-and-white stencil, the test had been administered before. The story was no longer than 106 lines. It was accompanied by a question sheet with ten questions. Most of the questions were on the outline, theme, characterisation, construction and title of the story. The short story and question sheet are part of appendix VI.

The second test was one of the regular tests on two units from the course materials *Unicom Finals*, which were administered every four of five weeks. The learners had worked on their own at unit 8. The knowledge and skills of the unit were subsequently tested in the following quinduum week. The test consisted of three parts. Part A focused on grammar and grammar/translation in five exercises of 5 or six items each. Part B was a letter presented in Dutch that had to be translated into English. It was focused on the grammar and vocab the learners had become acquainted with when doing the activities and task presented in unit 8. Part C was a reading comprehension test. It was a gapfill based on a text called *Online News Audience Becomes more Mainstream. Major newspaper sites lose audience share.* The learner had to choose one out of two options for the twelve gaps to fill.

The third test was again a literary test. It concerned a literary project in which groups of four read a novel on their own and reported on it. The test was a dossier test. This normally concerns activities that are not formally scored and graded. You can either have a pass or a fail. In this case, however, scores were given for the group essay.

Having made our first acquaintance with Mark, we will now turn to Pete.

7.5 Pete, the project man

In 1978 Pete started teaching two junior secondary classes for six periods a week. It was already at his present school. Pete was still a student at that time, with an MA-thesis to be finished. The six periods were gradually extended to 32 periods a week. From 1981-83, Pete also worked part-time at another secondary school. In the year of data collection, Pete was a full-time teacher of upper-secondary classes, i.e. forms 4, 5 and 6.

7.5.1 Formative experiences and core beliefs

Pete mentions three experiences that have proved to be formative to his teaching career: his cooperation with two progressive male colleagues, his two years' of additional teaching at the other secondary school, and the trust and teacher autonomy given to him by the school management.

First and foremost Pete stresses that he was greatly influenced by the views, beliefs and attitudes of two progressive colleagues, when he started teaching at his present school. The two male colleagues predominantly taught upper secondary education classes. They were in favour of new ideas, recent didactic approaches and progressive change. They had been actively involved in introducing and implementing the consequences of the Mammoth Act in 1968, which had completely restructured Dutch secondary education (See chapter 6)**.** Besides, the two former colleagues regularly participated in in-service teacher training programmes. Pete singles out one particular conviction his former colleagues held. Pete claims he was educated in the belief that text- and workbooks do not offer what teachers wish to bring into lessons.

> *I was still a little green. I'd just got my degree when six hours opened up at this school. And I thought, why don't I give it a try. I had just been living here and ended up in the presence of these two rather extraordinary gentlemen. These gentlemen were fully active in the circuit of 'refresher courses' and in those days our school was quite famous, infamous even, for the implementation of the Mammoth Act. They were*

quite involved with that down here. And these two people were quite progressive. And I was raised, if you will, by these two. So I was always brought up with the idea that textbooks can never offer that little bit of extra that you yourself can.
(3-1-2)

Pete adds that he is not sure whether without the coaching of his two colleagues he would have arrived at his core belief that materials developed by the teacher offer so much more than using regular text- and workbooks. Susceptible and inexperienced as he was as a beginning teacher, his colleagues' beliefs and practices gradually became his. Pete claims it still is one of his core beliefs.

A second professional belief of his is related to this first one. In view of the support he claimed he got from his colleagues, it is not surprising that Pete believes in collegial cooperation. His present working relationship with Joy echoes the relationship he once had with his two former colleagues.

P: // Yes, I'm not really sure whether I found that out myself, but in any case that's how I was raised and I still believe in it. I still believe that what you develop yourself contains more of yourself and therefore has more to do with the real person you are, and therefore comes across better in the classroom than some matter-of-fact book that you use year after year.

I: I see. And actually that insight comes from, as you mentioned, these two people who shared that view. You agreed with them and in the beginning you felt like, well //

P: // Well, of course I came when, well, I was so young. So you first feel like 'Well, let me just tone down a little and join in'. But after about two, three years I realized how true it all was. Lessons based on regular course materials are worse than lessons based on a teacher's own ideas and the materials he or she selects. It is just more fun that way. And that's the way we do things now. You saw that Joy and I were working here today. We really do basically everything together. The two of us prepare things together, discuss things together, we even evaluate things together. And what we don't like, we get rid of. And what actually worked, we file away. That's how an archive slowly began to take shape, an archive that's in a mess at the moment.
(3-1-4)

A second formative experience Pete mentions in his teaching and testing career was his part-time job at another secondary school from 1981-1983. Because he remained teaching at his present school, the number of periods taught by Pete in these two years totalled 32 per week. The headmaster of Pete's new school wished to modernise the English department, and thought appointing Pete would be a way to achieve this. Pete took up the challenge and accepted the post. It initially proved to be a period of hard confrontation. "Heavy, with lots of arguments and fights" (3-1-8). Pete would soon clash with one colleague and with one class in particular.

That headmaster really didn't like what was happening in the English department. And after the first skirmishes, I realised quite soon that if I were to push my way through, one of us would leave with a nervous breakdown. I mentioned this to the principal, but he just said 'Well, that's just too bad then'. He was tough as nails and I think that maybe I was, too. And it wasn't always fun.
That's also where I had my worst arguments with learners, with forms, which later turned into, and that's remarkable, into the greatest friendships. I was really quite young, you know, I was, I believe I had just turned 30. And when I was there, in the course of my first year, one 5 havo class just refused to be taught by me for two whole months. But when that was sorted out and they finally realised that my approach could also lead to good results. They really worked very well, and for the

first time in years their results were better than the national average. The rest of the school soon found out about this, and after that matters started to improve.
(3-1-8)

Pete was involved in appointing the new teacher who was to replace the person who had decided to quit her job. Pete claims this newly appointed colleague still enjoys working at this school. Gradually life was getting a little easier for Pete at this school. The learners and colleagues came to accept his different approach to English language teaching. Pete gives more details on his clash with the 5-havo class. Somewhere in November, the learners of this class told him they did not want to do what he wanted them to. Pete's response was immediate and drastic.

Yes, that was very embarrassing. No, it wasn't in the first weeks. It started somewhere in November. Then they said: 'We don't want to do what you want'. I said: 'Well, that's fine, then we won't do it'. At that moment, I just started to read my newspaper. At first, they gave me some funny looks, and then for a while they really// [two months (3-1-8)], for a while you'd see fewer and fewer kids entering the class, but at the beginning of each class I would ask them 'Do you feel like getting something done?' 'We don't ' Well, then I just opened my paper again.
They really got angry at a certain point and then I said 'Well guys, listen up. It doesn't matter to me whether I do something or not. I'm not the one taking exams. You are. And I've been hired to teach you and I decide what happens in this classroom, because that's just how undemocratic (laughter) I am. And if this is not what you want, you can go to the school management. 'We've already been there, but they told us that we have to talk to you about this.' I said: 'Well there's not much to talk about. I am willing to explain once more why I'm doing what I'm doing and why I think that's the right thing to do.' 'Well, why don't we get some work done?' I heard the first ones say. 'Have you lost it?', others reacted. Well, you know how it is. Anyway, at a certain point they came to me and mumbled 'Hmm, let's just do it.' Well and then I said 'Boy, then all of us really have to do some hard work. And we must all want the same thing, mustn't we?' 'Yeah, we must'.
And next they actually started and we moved all the desks around. I really can't stand desks being placed in three rows and all eyes being on me. And then they really went to work themselves and after that I even went to the pub with them. I treated them to some drinks, because we had had a good time and from that moment we actually developed a wonderful relationship.
(3-1-15)

Pete gives examples of the practices he objected to at his second school. He criticisms concerned invalid testing of listening skills, the absence of dictionaries and authentic materials in the language classroom and the way in which vocabulary was taught and tested.

How the approach at this school differed? Well, that's not very hard to tell. I'd better give you some example to illustrate this. The listening test, the easiest thing there is, the way they did it was by reading some literature out loud and then asking all kinds of difficult questions about it. And when I said 'Yes, well, but isn't this the simplest thing there is, because you can just get your materials from CITO, along with the national guidelines and then you know what to do and you just kind of prepare them for it. Well, they were flabbergasted. …. There were no dictionaries in the classrooms. They didn't subscribe to a newspaper. It was all just as traditional as German can be, you know. Having the learners study vocab, test it and that was the end of the matter. Testing wasn't even all that important. Well, you know, it was important to have some marks here and there and it kept the children busy. Studying lots of vocab.
(3-1-9)

A third formative experience Pete mentions is the trust and autonomy given to Pete and his colleagues by his present school management. He explains that his school management did not really play an active role in promoting teacher autonomy. Management allowed teachers to differ from one another in their didactic approaches and felt it was best to leave well alone. ("Laissez faire" 3-1-8). According to Pete, his approaches had been modern from the start of his career and are in line with what he thinks is expected of him in the second phase.

P: Well, I have to talk about the upper forms, because we've always used the regular course materials for the junior forms. And I think that's pretty good you know, for the little ones to have some form of structure. In the upper forms we've always done a lot of group work. We've always focused on a communicative approach. We've always worked on developing learner autonomy. They've always been quite independent here. Yes, and I think that then you're pretty much headed in the right direction, don't you?

I: Certainly. Working in groups, a communicative approach, learner autonomy and the independence // what was the difference between those last two things you mentioned? You said that learners have always been quite independent?

P: Well, you can interpret that in two ways, you know, for them to be in the classroom. But I think that over the years we've done a number of projects with our learners, for which they really didn't have to be in the classroom. And that wasn't mandatory at our school. They have a task, they have to complete it and I guide them through the process. And the more or less mandatory aspect of having to be in a classroom has never been the case here. (3-1-7)

I: So what you're saying is that your school managed cooperated with you on this.

P: Yes, yes, they never objected.

I: Consciously, or did you //

P: Nope, LAISSEZ FAIRE (laughter)

I : All right then.

P: English especially has always been kind of the odd subject out, you know, at this school.

I: Is it actually possible to be the odd subject out at this school?

P: Yes, yes, and that's really why I've always stayed here.

I: Do you also mean that in a negative way, that you say, without mentioning specific names or subjects, but that it also means that you //

P: // Yes, it's quite difficult to get everyone to agree on something, because everyone is allowed to do what they want. And that can be a bother at times. But on the other hand I sometimes think 'Well, I do it my way and friend b does it his way, never mind'. Also within our department we experienced two groups where the one group didn't want to join the other at all with what they wanted to do. But the same goes for that other group. You can best sum it up by saying that many roads lead to Rome.

(3-1-8)

Because Pete generally taught without using regular course materials, he was used to working with lesson series and projects, which were increasingly designed by himself. He was also responsible for the ways in which the knowledge and skills of the projects was assessed. The lesson series and projects themselves integrated such a variety of knowledge and skills that learners appeared to do well on proficiency tests, such as the CITO national listening test, with little test preparation.

Pete distinguishes between junior and upper secondary teaching. He welcomes published course materials in the first three forms of secondary education, because he feels young learners benefit from the structure they offer. When Pete starts teaching, he initially teaches forms 1 and 2. He sees designing and administering language tests in junior secondary classes as unproblematic, because the tests were directly related to what was offered in the course materials and to how this was done. However, he soon realised that reliable assessment of the same tests by various teachers was a problem area. When Pete came to teach upper secondary classes, he first used the tests his colleagues had designed. Gradually, he started constructing his own tests, which were largely modelled on the tests his two progressive colleagues considered effective.

> *P: As to testing, first I got those junior forms, because naturally I started in the junior forms. I had a first and a second form and testing wasn't a problem there, because we used the published course materials and the tests came along with it. Can I continue?*
>
> *I: Yes, please do.*
>
> *P: Then let me take a side track a little. Because I soon noticed that of course you can administer one and the same test for different classes. Yet, the ways in which this test is assessed can differ quite a lot.*
> *Where one teacher really wants you to dot all the Is and cross the Ts, well I often thought 'Oh, that one little letter really isn't all that important, as long as I get what it says there'.*
> *Anyway, then I got the upper forms and well, the first tests I used there were products of my colleagues and then for lack of better example I started making the same kinds of tests. And I even think I still use some of them in an ameliorated fashion, because they actually weren't all that bad given the time when they were first administered.*
> (3-1-2)

We have referred to Pete as 'the project man', and he appears to live up to that name. Working without regular text- and workbooks in upper-secondary education has stimulated Pete to design and develop lesson series and projects that illustrate his beliefs in effective teaching and testing. He claims that his projects contained a variety of authentic oral and written texts and tasks form the start, with abundant use of video and TV. Pete sees a direct link between the texts and tasks of his projects and the satisfactory results his learners achieve on the national CITO listening examinations[5], with little or no direct preparation.

> *P: I think that because at this school we work a lot with projects, in which the different skills are integrated and TV is used quite a lot. Well, it is a fact that we always do well on those (i.e., CITO listening proficiency tests) and I don't believe that my learners are better than other learners. I think they are somewhat better prepared, precisely because we do not exclusively use the CITO tests as practice materials.*
> *What happens here is that we just do our regular programme and then at a certain point we have the CITO test in 5 havo. And then about a month before the test, I say 'Let's just practise one'. Just to see where they stand. And then you select some learners who didn't do so well, and in that sense//*

that is required a little, isn't it, diagnostic assessments of this kind? So they do get to take home a specific programme to watch on TV. Something they can do, and well if they don't do it, well, they just don't. That's not my problem. And if they do, then you usually see their progress. Yes, and then during the// about two weeks later, we will do another one. And if I do three tests in total then that's quite enough.

I: Yes, yes. So in your case the CITO test is clearly a test, a moment of evaluation. The preparation for it//

P: // that happens in other ways.

I: For example in your projects.

P: Yes, yes

I: And you bring together the different skills in those projects?

P: Yes, yes.
(3-1-9)

Pete gives an example of one of his projects. In retrospect, it appeared to be a literary project that was be referred to in the course of the year of data collection. The project is called *Cal* and derives its name from a novel by the Irish novelist Bernard MacLaverty. The project explores the Irish question in ways that allegedly appeal both to Pete and to his adolescent learners.

Let me give you an example of one of my literary projects. A book that's been turned into a movie. So I combine sections of the book with the entire movie. Next, a number of assignments or exercises are added. Well then, if you're using a politically flavoured book, you could add all kinds of background information. You can then actually use taped materials, interviews, and whatsoever. The kids go on the Internet to find some extra things to go along with it. Well, we do things of this kind, and then the assignments, well, they're speaking exercises. They discuss maters, do some writing, do some watching and analyse fragments we viewed and things like that.
(3-1-10)

His approach to projects has developed over the years. His adolescent learners have more and more adjusted to what he refers to as 'zap work'. They are used to zapping from one fleeting highlight to another. In general, Pete feels that his present learners cannot concentrate on subjects and tasks as hard or as long as his past learners used to be able to.

At a certain point the projects you develop are adapted more to the learners. I used to create projects that could last up to eight, nine weeks, but if I now develop a project that lasts three weeks, I find that very long. I believe that my learners have become more fleeting in that sense, if I may say so. The span of interest, is becoming shorter and shorter. Because of the whole business with the Internet and also television, everything has become like zap work and I've adjusted myself to that a bit. It's a little less deep at times. But on the other hand, maybe the pace has gone up a bit. Because they keep getting new things thrown at them, the span of interest does remain taut and I believe they do take in a number of things. And maybe it has become a little easier than it used to be. I really used to be able to spend two months on, well, shall I give you an example? Well, John Lennon died, you know. And I believe I had prepared this project in January and February and I think that project lasted for the remainder of the year. The kids really quite liked it back then. But I wouldn't ever be able to do that these days.
(3-1-11)

Yet, Pete does not easily give in to the learners' tendency to zap from one point of interest to another. He tries to keep the learners focused. Language learning motivation is a key issue for Pete in selecting the subject matter for his projects.

> *I feel that it should interest my learners. It doesn't all have to be as easy and fleeting. The subjects could also be political. But it does have to be presented in a way that it's kind of pushed into their direction. It does me no good whatsoever to deal with Shakespeare sonnets in 4-havo. Because then I am doing all of the work, whereas really I shouldn't. It does do me some good to show 'Shakespeare in love' to 5-havo and then to talk about it with them, like 'Gosh, how do you feel about that dude walking through life so happily'. You can do that, and I use the word 'dude' on purpose here. So it doesn't become so high and mighty and vague that the learners can't use any of it at all, is my belief. But there are others who totally disagree with me. …. When I talk about it to colleagues, then I tell them 'You're all just performing in a play, you and your learners'. Because your learners don't understand a thing about the original text. They don't even like it. They just look for a summary and for the oral examination they just rehash that summary and you say 'Well, isn't that nice'. But the only thing that's happened is that they've studied that summary. And you know that, too, and then you're just talking about nothingness together. That's some kind of game, which I don't go along with. And I don't want to. I can only talk to learners about books that I know, so I'm limiting their reading lists. And they understand, I do always have to explain it, but they understand. But I can't ask questions about a book that I've only read the summary of. Of course, I easily could, but I can't attach a mark to that. Attach an assessment to it, which I find pretty difficult anyway, compared to asking a learner about a book that I do know, and which maybe I even loved.*
> (3-1-11)

Pete's beliefs are reflected in the way in which he designs his projects. He feels that the subject matter has to engage the learners from the start. Projects may be introduced by an appealing text, documentary, song or poem. In the projects Pete integrates the four linguistic skills with what he refers to as 'vague skills', such as self-regulated learning and activities and tasks that involve real communication.

> *The design remains the same. You intertwine the four linguistic skills with a number of vague skills, such as self-directed learning or learning how to communicate. And that remains the focus and anything could give rise to a project. It could be a good text, an attractive documentary, some heavy song I heard on the radio, a beautiful poem, or// . You name it, and we can turn it into a project (laughter).*
> (3-1-11)

Pete acknowledges that the learners attach importance to the tests related to the particular knowledge and skills dealt with in a project. He realises that the learners depend on the scores they get. That is why tests are generally planned before the start of a project. Pete feels it is important that the tasks and activities that precede the actual test enable the learners to achieve good results.

> *P: Well, I think that you should first plan your tests for the year. Once again, that's what the projects have to be related to. So I find that the most important thing. That is to say, I don't find it important, but it's really vital to those kids. Marks really are important to them. Yet marks do not include assessments of effort or anything of the kind. It's an assessment of 'what you can do' or 'what you might be able to learn'. So first we plan the tests, then what kind of tests they are going to be and then what we're going to plan around it. Therefore whatever I do is in preparation of a test. If I give a grammar test, then I really find that a preparation for later writing tests, but the period preceding the test will be devoted to grammar.*
> (3-1-12)

Pete mentions a building metaphor to explain how his tests generally move from discrete point grammar tests to integrative tests of more complex skills.

> *We actually start with rather simple things. And they, in turn, become bigger and bigger tests. So we'll start off with a grammar test, but I really find that a preparation for later writing tests. Because that grammar consists of little building blocks, together with the vocabulary in your own brain, with which you're going to work later, with which you can express your thoughts. And yes, you can't build a house, I always say, unless you know at least the tiniest thing about placing the blocks together.*
> *So you have to be able to convey the message, and to do so you need the proper tools. And in that sense I also feel that they should use the dictionary sparingly. I am not keen on haphazardly finding a really difficult word in a very simple sentence, because I find that out of keeping.*
> (3-1-24)

A final formative experience in his language testing practice has to do with experience and age. Pete feels he has gradually past his 'angry young man' stage, and has become milder and more supportive in his assessments. This is particularly the case when complex skills have to be assessed or with assessments that are likely to be subjective.

> *Earlier you asked if there was any development in my assessment practices. I think that surfaces more and more often. The older you get, the milder you become. I used to be some kind of an 'angry young man' here at school. I think I have become milder over the years and that I approach matters more positively. You can also see that if you look at the marks I hand out, and especially the subjective marks. I think that a learner who so very much does his best, but in the end isn't very good, I still give him a nice mark. Because it stimulates him to continue. And I know from experience that as long as they just continue, they'll eventually turn out all right. So they will also get a six for that multiple choice test. And if they don't succeed, I'm decent enough to make sure they're going to have a six average for their school examinations, and then we'll just hope that they'll get a five, so they'll have a six as their final mark and then we're both very pleased. For me to put the skill called multiple choice on a pedestal, just because it's so easy to score, for me that's just an easy way out.*
> (3-1-19)

Another of Pete's core beliefs concerns effective grammar and idiom teaching and testing. Pete detests grammar tests in which the learners have to mechanically reproduce tricks, so that the correct forms are produced. In mastering English verbal tense and form Pete acknowledges that learners may benefit from a simple and straightforward heuristic scheme. The scheme is presented to the learners and they practice with it. To test the learners' knowledge of the English verb, Pete always starts with an authentic article on a subject he expects to be of interest to most of his learners. He then deletes a number of verb forms. The infinitives of the omitted verb forms are then given. The learners have to put the infinitives into appropriate tenses and forms, with or without the help of the heuristic scheme.

Pete has similar views on idiom testing. Authentic texts are the starting point of his idiom tests as well. Vocabulary is preferably tested by leaving out items in some attractive story line.

> *Grammar consists of verbs for about 90%. What I often do is using newspaper articles. I delete the verb forms and provide the infinitives at the bottom. And then they have to put them into the correct forms n the right places in the article. And they have to do that according to a system I have taught them. And they have to be able to come up with the right form in line with the contents of the text. Or an idiom test, have you ever seen one of Joy's? Well, we often prepare// Those idiom books are really all*

the same. There are these groups of words listed, supposedly according to subject. But I often fail to understand the relevance of the subjects the idioms relate to, but anyway. But next I'll use the idioms the learners had to study to concoct some weird kind of story. Something a bit exciting, a bit bloody, vampires and crime and such. And kids will read that story and have to fill in the words, which are listed in alphabetical order, with a couple of words stuck in there that don't make any sense at all. And so they have to combine the content of the story with the meaning of those words. To make some sense of it.
(3-1-20)

Pete holds a passionate plea against idiom tests in which the learners reproduce words and phrases that lack any meaningful context.

P: Sometimes, yes. But I don't just hand it to them on a silver platter, because those tests can really be terrifyingly difficult and I'm not the kind of person who teaches them tricks and they really have to get used to that. They find it very difficult in the beginning. Later, they're getting better and better.

I: What do you consider to be a trick? Maybe that's important to…

P: Well, I study an abstract word that I, for instance in Dutch I don't really know what a 'bahco' is.
I know it has something to do with tools, but I really don't know what it looks like. But I could put an English word next to that and just study that, but then I haven't really learned anything at all.
I really don't know what to do with it. That's what I call teaching them tricks. Teaching them tricks, that's also when you give them a text, tell them to study all the words in the text and then I'll quiz them on a number of those words, yes. Does that do any good? That's what I always ask myself, does it do any good? Well, do we say that it does, well, I am not even sure of that. Well, so then I think don't do it .

I: When is an idiom test useful, to you?

P: …When I make them study words where I know that they're important in the upcoming multiple choice final examination. And you JUST CAN'T DO IT (laughter). Because that's something you never know. And in that sense I find it more important that they're able to guess, that they can guess the meaning of a word. And I do a lot with that, and I do a lot with making that connection. Kids who take French really don't understand, for example, that eight out of ten difficult English words are actually French. And I tell them that, and then they think about that. I'll say 'Well, if that is what it says, what would that difficult word mean?' Now you don't need a dictionary for that. To teach them that. ..
But my own son studies idiom and he says oh well and all right and then he makes an eight or a nine. And then I think well what on earth is that good for? For nothing, isn't it?. And it's a very time-consuming task for most learners. It's, it's// Yes, it shouldn't be like that. It shouldn't be like that. So I think an idiom test is a very bad. It's the worst possible type of testing.

I: Yes, yes. So according to you, an effective written test won't have any reproduction of idioms?

P: That's right.
(3-1-21)

In general Pete feels that discrete-point tests depend on applying pre-concocted tips and tricks. The results of these tests are unrelated to a foreign language learner's proficiency. Therefore, he prefers to work towards larger integrative tests.

P: Well, I prefer tests that take more than those little fumbled things. I don't do those discrete-point tests, because their nature often lies in the trick-teaching trade.

I: Yes. So you'd rather work towards a larger integrative test after a certain amount of time?

P: Yes, indeed.
(3-1-24)

Halfway the first interview Pete summarises his main language testing concerns.

I: You've already mentioned that a lot, I think, Pete //

P: //but not really regarding testing, I think, or did I (laughter) ? Maybe a little, maybe a little, yes.

I: What aspects did you mention?

P: (With deliberation) That the lesson content and the tests have to be related. That tests are important in the sense that they're decisive for the learners' advancement into the next school year. And that I often find the assessment difficult, because it remains subjective and that I'm looking for objectivity within my own subjectivity, by forcing myself to look back at older stuff.
(3-1-16)

In teaching and testing Pete objects to 'vagueness'. He uses the adjective vague in relation to the constructs learner autonomy and communicative language education and in relation to the skills of speaking and writing. Pete struggles with straightforward assessment criteria for the productive skills. In that sense, using the national CITO listening and reading comprehension tests is a relief.

P: Yes, well. A very important aspect and at the same time the most difficult thing is that a test should really be objective. But that's the thing I find so difficult to accomplish with those linguistic skills. Then I really have to twist and turn in all kinds of strange moves. I can do it with a listening test and the multiple choice test, because there, there I manage to be objective. But I don't succeed with writing and speaking. There I have to find my own objectivity, but in a desperate way. That the entire time I'm like, phew, am I doing this right? Is this good?

I: With regard to writing at least you gave that example of the way you deal with//

P: Yes.
(3-1-23)

In view of reliable assessment, Pete is pleased with the educational packages for listening skills offered by CITO. They are of help, and so are the national tests they offer. Pete feels that the tests reliably supply information on how your learners are doing in comparison with others. Yet, he is also critical of the way in which listening skills are assessed. Pete feels it is fairer and more valid to test learners by using video. TV has replaced radio as an important medium for learners.

I: One aspect of testing you just mentioned, didn't you, related to listening. That you say that there are national guidelines. Apparently it is of some use when you start teaching listening skills.

P: Yes, it does in so far as I think my learners are only here for one thing, which is to get their certificate. And for you as a teacher it's important to help them on the right path and anything that's 'national' does provide something to go by. And in that sense I also use those CITO tests, even though, if I'm being really honest, I don't always find them that great. I much prefer working with video cassettes, because listening // look, I think that CITO listening tests were constructed in, what shall we say, the late sixties? I don't even really know, but that's what I guess. And in those days, kids still listened a lot to the radio, I think. But television has replaced the radio as the main medium for listening. And that's why I think you should just go along with that.
(3-1-9)

Pete is also critical of the way in which CITO measures reading comprehension. He questions the validity of the way in which foreign language reading comprehension is tested in the national examinations.

P: I also often find that multiple choice texts appeal more to intelligence than to knowledge.

I: General knowledge, that's what you're talking about, advance knowledge.

P: Yes, yes.
(3-1-28)

However, Pete is most critical of the training and testing packages CITO offers for assessing foreign language writing skills. He does not consider these very helpful. The letters are pre-structured and constructed by some testing expert. Pete misses the learner's own voice, a voice he would like his learners to develop in their years at secondary school.

P: But with the somewhat vaguer things I do take up a more vulnerable position by doing it that way. And if people say 'Here's a letter, I didn't write the letter, but I put the information in chronological order and now you're supposed to, well, let's see, well I believe the date should be in the upper right hand corner, I'm not even sure. And if you start giving them marks for that kind of technical stuff, sometimes I think 'Well, that's all nice and well and then you can say that I've taught them a trick, but what's really the point?'
(3-1-19)

A past experience with applying CITO assessment criteria on English writing of secondary learners has strengthened Pete's belief that the productive skills cannot but be assessed holistically. This is what had once happened.

P: I remember very well when CITO started constructing those writing tests. I once went to one of those introductory days, together with one of the two colleagues I was talking about in the beginning. There was this system with circles and plusses and minuses and then a circle signified a grammatical error and the plus a writing error. I'm just making my point here. And then you had to check it all like that and we looked at it. And, then, well we read that letter and I said to my colleague 'Well, it looks like it's between a three and a four'. And my colleague says 'Gosh, I was thinking the same. What shall we put down, a 3.5?' And after just reading it once, we had finished already and we felt really silly that we hadn't understood the system of scoring very well. But it turned out that the actual score was in fact around a 4. And then we

just said 'Well, let's forget about the scoring system, because it has no added value. It's just that you're creating some kind of semi-equality'. But when I have a kid …. write something about something, say something about the homeless or being an addict, or something they've experienced themselves, then you can't just apply that system as such. It becomes subjective. Because one kid can put it in better words than another and I don't really mind that and it doesn't have to lead to unsatisfactory marks. I sometimes say that if you're able to make sentences with four, five words and you do that first without making mistakes, then you can go to the next level and try to unite some of the sentences, so there's some variation in the length of the sentences. And then everything will be all right and it doesn't all have to be that brilliant to be satisfactory. Well, and that's the way we try to do it.
(3-1-12)

Pete feels that productive skills cannot be but assessed holistically. As a teacher you are invariably led by your knowledge of the type of learner and the effort and emotion (s)he has put in a test task. Yet, Pete acknowledges that some tests are easier to assess than others. He prefers to start with tests that are more "learnable" and easier to assess.

I: Does testing play a part in that process of gradually moving away from the classroom system //

P: // Yes. Tests are more learnable in the beginning than towards the end.

I: Can you give an example of that? A concrete//

P: Well, a very simple grammar test is learnable. You can test the verb forms, no matter how difficult they may be for some students, but there's a system to it that I explain to them and they can work with that.
And I make my tests in such a way that they fit in with that system. Because otherwise we're doing the wrong thing here. But in the end we get to the writing tests and then you enter a somewhat vaguer area.
(3-1-12)

When Pete assesses writing skills, he claims his holistic assessment procedure is often as follows:

P: // Yes, when I assess a writing test, first I always thumb through some old tests of the kids of some years before. What did I do here and well, then I kind of remember how it was and only then do I begin. It really wears you down. It's the only examination that really takes a lot of time to check. I hardly ever write anything on the papers in the beginning, so they're quite immaculate at first, but I do order them in little stacks. So then I think 8, 7, 6, 5, 4, it should be something like that.
And then I start comparing the two stacks and I'll take a closer look to see, 'well, how's the spelling and why did I like this so much?'. Yes, each of them is read about four to five times before it finally receives a mark, which is made up of a combination of spelling, grammar and content, where, because it is writing, the grammar will be weighted more. Because I myself am quite sloppy when spelling is concerned, spelling errors are weighted less. I'm not going to weigh something that I'm not very good at so much in my final assessment. Besides, when they do this on the computer, half of the spelling errors will be gone when they use a good spelling checker.
(3-1-13)

Pete feels his preferred assessment procedure is justified because his learners never complain.

P: It's remarkable that my learners have never come to me to complain about their marks. Never ever. Well, they do if I've made a mistake while checking the test, but with the vague tests I've never had any acceptance problems regarding marks, whereas other teachers meet with problems all the time. My own sons can go pretty far with some teachers. They'll go up to them even if they can't find a mistake and then they'll say 'But sir, isn't this really quite all right'. And then that teacher will read the answer again and think well why not, and adds 0.5 to the final score. And why he's actually giving the 0.5 nobody can tell. I always tell them 'Guys, you can always come complain to me, but I don't think it will get anywhere.' But then again, they never do.
(3-1-16)

Pete's concerns with reliable and valid assessment criteria also relates to what is commonly referred to as 'native-speaker level'. Pete feels there is no such thing as native-speaker level. He is quite adamant about it on a number of occasions. Irritated by efforts to try and turn learners into near-native speakers, Pete defies the construct.

P: And we as teachers often have the tendency to try to approach native-speaker level. Well, that's just what you will never attain, so you can't expect that of the average high school learner either. It is even stated in the second phase attainment targets that vwo-learners should be able to approach native-speaker level. Well, that's just wishful thinking, because a: those kids are much too young for that, and b: 80% of them have never been to England, so where should they acquire this level? From MTV talk? Is that near-native speaker? When are you a near-native speaker? I watched Boys in the hood this week and it's all about 'Shit man' and 'fucking' this and that, and well, my son can say that. So is he a near-native speaker? But those black guys in the movie were more or less without any brain content, except for those two that the movie was about. But they're all native speakers. But they don't really say anything. So what's that then?
(3-1-22)

Later on in the interview, Pete's irritation with native-speaker level returns.

P: Tell me now, what's that// it says that a child should be able to do more of this and that at a native speaker level. But it starts out on the wrong foot by saying that it should be more 'native speaker', because what is that? Which native speaker? John Doe watching football, and if he turns out wrong, he's beating up the cops or blacks when he's in his early twenties? That's another difference by the way, is he white or black? Where's he from? What is that?

I: Yes, that's very clear, yes.

P: Usually it's someone who's passed his four A-levels with flying colours. Because there are also English people who have a vocabulary of no more than 2,000 words. And they are still native speakers.

I: Just go to the Spanish Costas for a holiday. [Peter laughs]

P: Yes, by the way I have just read this great book about football hooligans, who really do use the 'f-word' every two words, you know.
(3-1-27)

Pete realises that compulsory reading is a struggle for most of his learners. He is willing to make life easier for them, if possible.

P: But, just try to give a mark. When learners intend to read English books, I usually tell them: "Just read that book. Select one, and read about 15 pages. And If you think, well, it's going to be a nice book, read all of it. And 'nice' in itself already is a very subjective term, because, and I realise that, reading takes away a lot of your free time. But anyway, we just have to get over that, and so you've got that book, and you like it. So do write a brochure about it, in which you pay attention to so and so.
(3-1-19)

Pete is of opinion that colleagues who work in junior and upper secondary education should be cooperating more closely. Pete regrets there is no real dialogue and agreement on important didactic principles is often lacking. Pete and Joy are employed at a large comprehensive with several locations. Forms 1-3 of havo and vwo education are situated at another location. This makes the problem all the more poignant.

P: I think that the main defect of the second phase is that all those schools have become so large that the junior forms actually are in a different location. It's the situation not only at this school, but at a lot of other schools as well. So information gaps occur between the different teachers there. I have no idea what's happening in the junior forms. I hardly even know what's happening in the third forms. And I believe that the bridge between the third and fourth form should have a much sounder foundation than is the case at this time. But what happened is when the second phase came, degree-one teachers threw themselves at it and the degree-two teachers kind of thought 'Well, it's nothing to do with us'. At least, that's what happened at this school. And I think this has happened at a lot of schools.

I: Like, we're just going to continue doing what we've always done. With grammar tests//

P: // Yes, and I've always said// Many years ago, I used to have arguments about it at this school. What is actually going on? A havo-level learner is going to have his final exams. He's going to have them in five years, and that's why we do this in the fifth form, and that's why we do that in the fourth, and that's why so-and-so has to be done in the third and this in the second and that in the first. Well, it doesn't work like that. It doesn't happen like that. Reality is very different. But that's the way it should be.
(3-1-19)

Pete really believes that all of the years of secondary education should link up well. To achieve this end, degree-two and degree-one teachers should be cooperating more closely.

Pete's classroom management is characterised by a personal interest in the well-being of his learners. He stresses the importance of atmosphere and learners expressing what they stand for when fluency is tested in groups.

P: The weather, your hobbies, a subject you'd like to discuss. I believe Joy has told you how we do that these days.

I: She has, indeed.

P: About groups, and what they can do. You often experience hilarious things. It's often a lot of fun, and I believe that my learners appreciate it. That with me an oral examination can actually be fun. They find it a bit odd. It's all so relaxed. But isn't that great? Shouldn't it be that way? We're actually saying goodbye a bit at this stage, and we should end with good feelings.

I: Yes, yes, definitely. In fact you're saying 'Gosh, sometimes it's just hilarious, fun, personal. It's meaningful'.

P: They find it quite difficult actually, to talk about themselves. But it does happen, at times. There's no need to do that all the time, of course, but it does happen.
(3-1-13)

It is important for Pete to gradually get to know his learners. He tries not to rush them, though realising that most them are not at school out of their own free will. The didactic procedure of writing a letter of introduction Pete refers, was also one of Mark's favourites.

P: Yes. Yes. Last week I heard something like that on the radio… It was this workshop thingy for young people who were getting married. To teach them what it actually meant. And shouldn't you do something like that in secondary school and then the man said: 'Yes, you have to be careful, because those learners aren't there out of their own free will. They're stuffed into some kind of class and you're put in with them and you have to slowly start creating trust.' But I do demand something personal. I'm going to write to my learners, I'm going to do this tonight, an introductory note, telling them something personal about me; where I live, what I do, what my hobbies are, what my family looks like. And then they get to take that home in an envelope, which they don't get to open until they're home. They'll probably open it in the school yard, because it's very exciting of course and I do expect them to write something about themselves too.
(3-1-14)

In his teaching and testing practice, Pete primarily focuses on the motivated learner who works hard, but is unsuccessful. He is prepared to motivate them by giving them marks that are higher than they would objectively be entitled to. Learners who underachieve are told off once or twice, but are basically left to themselves.

I: Do you feel that learners become more skilled in languages if you reward the effort they have put into their work?

P: // Yes. They are prepared to take greater risks, because they … of course there aren't that many who actually do that, you know. Most learners just kind of manage to scrape through and then you have some good ones who can do a lot by doing nothing and then you've got a couple that have a lot of trouble getting by. And you get that with English because, everybody takes it, just like Dutch. I think that's a comparable subject. It's vague too. (Both laugh)

I: Still, you've managed to say a lot of concrete things about such a vague subject.

P: Yes, I have, yes. [I'm sure of it] And I think about it every day. What I wanted to say was / I don't remember what I wanted to say. We were talking about, about this vagueness that .. that it's stimulating, right. Incomprehensible //

I: // We established the link with Dutch, didn't we?

P: // Yes, I think it's very good if a learner is able to do more and more, while a couple of learners are still struggling with it. Oh yes, that's what I wanted to say. You have some who really have great trouble, and I help them, I don't help the entire pack, because most of them can manage. They understand that you're helping those select few, and you can neatly focus on them. Because that's what it's all about.

I find the learner who scores sixes, whereas he is capable of scoring sevens, a real dork. And I am stating it mildly, because that tape is running (laughter). But I really find him a proper nitwit, because he's selling himself short and I do tell him that as well. And then I'll mention the marks. I tell him when you make a seven, you'll have an extra mark Ito compensate and who knows, you might get this math problem that you can't possibly get right and then you might need that seven, and so then you're a major dunce if don't try to make that seven in my class. 'Yes, but I just don't feel like working.' – 'Well, you don't have to, it's not my business'. And it isn't. But kids who are really giving it their all to make the best of it move me, and they're the ones you've got to help. And I really believe that I do help them. I would find myself a really lousy teacher if I didn't do anything for them. So I do, yes, yes, all right.
(3-1-18)

Before turning to Pete's construct definitions, here is how he looks ahead at the first year of the curricular and didactic innovations of the first year. Pete approaches the innovations of the second phase in a realistic way. He is a survivor, sticking firmly to his beliefs in the seas of change, willing to float with the tide, yet unafraid to sail upstream if need be.

I: So now you're mentioning the organisational aspect of the second phase, aren't you. The fact that you're going to work with profiles, that the number of subjects is growing and of course that both the teachers and the learners are going to be under a lot of pressure this coming year. ..

P: Yes, but I think that maybe the first and the second year will be a bit problematic, but that will take care of itself.

I: Because education has weathered more storms in the past.

P: Yes, it has.
(3-1-19)

In spite of Pete's dislike of working with course materials, Pete and Joy had decided to partly work with a course book in the first official year of the *second phase*. In view of the changes to come, they both felt it was better to be safe than sorry. They opted for buying classroom copies of *Touchdown* to help them out with administrative matters and make use of learner materials such as evaluation grids.

I: Do you have any experience with course materials, a detached course book, as you call it, where you thought like, well, I tried it but it's all wrong?

P: Happened hundreds of times. Of course, because parents and learners sometimes call for course books. It just happened today, in fact. Because of this second phase, there are so many administrative matters to take care of. So much has to be written down that I really felt last year that, well, I'd better rest my head in my lap for a bit because there's no way that I'm going to be able to get this done on my own. So in the end we opted for Touchdown. But not every learner's getting it. We're just going to put several copies in the classroom, and before the holidays we agreed to do some things from the first few chapters. Joy was going to prepare that and yes, today she said once again " Well, we'll just have to do it." But I'm not sure.

I: So, in a way, the reason you bought Touchdown was mainly to please the parents?

P: No, I think we have to meet the ministry's demands at this time. To make all those evaluation forms yourself is just, and then if something's wrong at least you're covered. (3-1-6)

This is where our first introduction to Pete ends. His beliefs are strongly related to two formative experiences. Pete was still an MA-student when he started his teaching career. Two progressive colleagues set examples, which he was to follow up and further develop in the rest of his career. One of Pete's core beliefs is that course materials do not bring what self-developed materials, projects and tests do. Another important formative experience was his part-time job as an English teacher at another school. Here he was hired to introduce some fresh ideas to the teachers working in the English department. His ideas were not accepted by one colleague, who left school, and initially by one 5-havo class.

In the next sections, we learn more about Pete's interpretations of learner autonomy, communicative language education and foreign language assessment and evaluation.

7.5.2 Learner autonomy

Pete first of all expresses the need for a gradual approach towards learner autonomy.

P: Yes they come straight from the traditional classroom system and then they're pretty much thrown into the deep end with me. And I've learned that if you do that too quickly, you will create a gap and pretty soon some learners are going to disengage, because they just don't understand what I'd like them to do, nor do they get what they should be doing themselves. So I try to build it up quite carefully. And then the autonomy increases more and more.

I: And testing plays a part in that process of gradually moving away from the class system //

P: // Yes. Tests are more learnable in the beginning than they are towards the end.
(3-1-29)

Pete tells us that no person has ever succeeded in telling him about the differences between self-regulated learning and learning with full learner responsibility. In defining self-regulated learning, Pete refers to tandem and group learning.

I: Self-regulated learning and learning with full learner responsibility, the things everybody is talking about//

P: // Yes, the latter seems to be a step further. But nobody has ever really clarified that term to me either.
But self-regulated learning is indeed one learner consulting another. I really do believe that when you put a learner who really understands that system of the verbs, for instance, to name an easy thing, next to one who doesn't get it at all, and they're both quite willing to make something of it. Then, the learner who doesn't get it, will learn much more from the good learner than he will ever learn from me, because the distance between us is too great in that respect.

I: You tend to leave out steps?

P: I'm just guessing. I don't do it on purpose, because I do sometimes try to play the fool. I just think that they'll learn more from each other than they will from learning on their own. And I truly believe that.

I: By which you're saying?

P: In that sense, a, a, and it so happens that I think English is a hands-on subject, for which you don't have to learn many tricks. Because I always say, well, if you ever want to understand that text, you're going to have to understand about 70% of the words in that text. And then they have to be, and then we're leaving out the articles, right, they have to be pretty reasonable words. And then you have to develop some kind of skill, with which you activate the part of your brain that does the guessing, so that you can still place the words that you think you really can't possibly know in the right context. And that's just a matter of practising. And in that sense I believe that it's better to look up a word like that ten times than to study it from some idiom book just once. Because why are you studying that word, you're studying it to get a good mark. But the mark really isn't important. It's the knowledge that really is.

I: Yes, I definitely. By using it or seeing it a lot, you actually learn a word.

P: Yes, yes. And not just by looking it up, but also by writing it down.

I: Definitely.
(3-1-29)

We attempt to find out in what ways learning with full learner responsibility may differ from self-regulated learning. Pete seems to feel that his learners are not of the right age to have a say in what has to be learned. He feels adolescents are not up to that task, unless they have been given the opportunity from the first form onwards.

I: Learning with full learner responsibility is a step further, you mentioned. In what way is this the case?

P: Well, because then learners start to make decisions about the content and I wonder whether that's actually possible. ….. So then they start deciding which tasks they'll do. I don't think a child is capable of doing that. He could, but then you have to let him do that from the first form onwards, and not suddenly when they reach adolescence.

I: So if you start doing that in the fourth form, really you're too late?

P: Yes, then you are way too late.
(3-1-30)

We probe a little further and ask whether discussing learning and learning processes with the learners would make them more responsible for their own learning. Pete seems to warm up a little to the idea and responds with a tentative maybe.

I: Learning with full learner responsibility, does that have anything to do with teachers expecting their learners to speak up? That they can formulate a point of view? Or isn't that an aspect of responsibility on the learner's part?

P: I don't know, I'm not so sure it is. If I ask a learner 'Do you think that what we did was useful?', would that be some form of learner responsibility?

I: If that is how you wish to interpret it.

P: Maybe it is, because if he really put a lot of effort into it and he found it a total waste of time, then you can ask the following question: "Then why on earth did you do it?"; " Why didn't you come to me sooner?"; and "Why didn't you ask me what the use was, because maybe then you would have understood or maybe I would have changed my mind?". And if the answer is it could have been useful, but I just didn't try. Then you've actually learned something. But is that learning with a sense of responsibility? Yes, I think it is, a little bit, maybe.

(3-1-31)

Even though Pete goes for a tentative maybe, he believes that there are limitations to the responsibility you can give to adolescent learners and that a gradual start is required.

7.5.3 Communicative language education

Some of Pete's core beliefs are reflected in his definition of communicative language education.

I: Yes, yes. What's communicative language education? Well, you already got it//

P: // Well, that you've got a message I'd like to know about, and that this message is understood in one way or the other.

I: Any way at all?

P: Yes, yes, that doesn't matter.

I: All right, that's clear and //

P: // and it should also have some kind of meaning. That's what I find lacking in the course books. They're often about nothingness. They're contrived.

I: Any examples?

P: Phew, Well, what's a seventeen year-old kid going to do with a text about Victorian philosophers? Then you should tell me how that could ever have been a text for the finals. .. Or what should a kid know about druids? I find that nothingness. That's a very difficult text, and that's all it will ever be. And a 17, 18 year-old, not at any level, is able to relate to that.

I: Because it'll never, or hardly ever, be used at a communicative level?

P: Right, right.

I: Yes, yes, all right.

P: I call that a why-should-I-care text. And it doesn't do anybody any good. You're not testing anything by using that.

I: So that's not what communicative language education is about, that's quite clear//

P: //No, no. I often find that multiple choice texts demand more intelligence than actual knowledge.

I: General knowledge you mean, previous knowledge.

P: Yes, yes.
(3-1-29)

We asked for a definition, and a definition is what we got. Pete feels that at the centre of communicative language education is that one person has a meaningful message another person wishes to hear and does understand one way or the other. He stresses that the message should be meaningful. They should be of interest to the interlocutors, which means that it ought to be about a subject they can relate to.

7.5.4 Foreign language assessment and evaluation

Below, we will report on Pete's views on assessment and evaluation. We will again report on the three question categories of the first interviews held with the three teacher respondents.

Essential knowledge, skills and understandings

Pete mentions discrete elements of knowledge as well as particular microskills and understandings that make up language proficiency.

I: The guide mentions that we would also talk about knowledge, insight and skills. So what exactly is knowledge of English? What is absolutely essential for a learner to know?

P: … That in English, the subject always precedes the verb. .. You shouldn't translate it literally. That you should try to say what you want or write what you want, and that you can convey a message with your knowledge of English. And then I will try really hard to forget I am Dutch when I assess and ask myself: 'Would an Englishman understand this?'. Is this gibberish, which I actually get a lot at oral exams? (laughter). Is it still understandable to an Englishman? Well, okay then. But sometimes I just don't get it. Then this learner is yakking away for ten minutes, and I'm yakking along with him, but then we're just not making sense. And that's when it really goes wrong. So, in that sense knowledge signifies that the learner should have some insight into how an Englishman will read this. He doesn't have to become an Englishman, because he's Dutch, but he should be able to convey the message. He should be able to, quite simply, show someone the way. When an Englishman asks, well, what do you think about that book, then he should be able to tell him about it a bit.

I: … Communication [yes] in a foreign language, yes.

P: And he needs a couple of skills to do so, yes. Very particular skills. Such as vocabulary, grammatical insight, and some insight as to who am I actually talking to, I find that important too. So, say, I'm writing a letter to a pal of mine, then I'm definitely going to use different language than when I'm, say, writing a letter to the school principal.

I: Formal, informal.

P: Yes, yes, yes. There's a certain difference there yes, and, and … (5s)
(3-1-25)

Pete enumerates several knowledge areas, which he gradually links to insight. The first is very concrete. Learners should know that the subject precedes the verb in English. Later in the segment, idiomatic knowledge and a summarising 'grammatical insight' are added to the concrete knowledge of the subject preceding the verb. This brings him to the learner's knowledge that one cannot just literally translate the

structure of Dutch into English. This knowledge is supposed to lead to better understanding of or insight in the language. A third knowledge category is that the learner realises that effort and perseverance are required to express what (s)he wants to express, orally and in writing. Learners should be willing to give effective communication a serious try at the least. A fourth category is the learner's awareness that (s)he can actually convey the message (s)he wished to convey. A final category relates to sociocultural knowledge, and more particularly knowledge of interpersonal relations related to class and/or power.

The united effort of knowledge areas leading to insight and understanding already preludes how Pete views insight. He philosophises on the complexity of the language proficiency construct.

> *P: // They're building blocks, aren't they, and you have to do something with them. You have to organise your thoughts with these building blocks in such a way that you can convey the message.*
>
> *I: Is that to you the insight phase?*
>
> *P: Yes, I think so, yes//*
>
> *I: // that you say with those language aspects you have to communicate, through listening, speaking, and writing.*
>
> *P: And it's also true in the passive sense.*
>
> *I: Yes, yes, OK.*
>
> *P: Is that an answer you can live with?*
>
> *I: Yes, yes, absolutely, I am happy to say I can live with all kinds of answers.*
>
> *P: Yes, yes. These are kind of like the vague things, you know, because when I prepare my classes, I do put those tests in place you know and then, and then I think, well, phew, phew, is there any use? What do I want to actually happen? Does it make sense or should I question what I'm doing as a teacher? I find that very important and well I'm never very insecure about what I'm capable of or about what I'm doing. I can present matters in such a way that everyone starts thinking 'Well, that must be really important'. But sometimes I think, 'Well, but what did we really do just now?' But in the end it always turns out for the best.*
>
> (3-1-26)

Pete hints at the important role of the teacher in motivating learners for a particular tasks. Experienced as he is, he is capable of engaging his learners even when he is not completely convinced of the use of what he is asking the learners to do.

Pete is asked whether he has ever attempted to address or to investigate matters that he refers to as "vague". He again replies that often there is no reason to, because the learning results are fine. Pete discloses that he sometimes turns to activities or tasks that need no teacher preparation, such as having the learners do MC reading comprehension texts. Much to his surprise, his learners appear to appreciate the change for a while.

> *I: Doesn't that lead you to change matters, when you say well, what did we really do just now?*

P: Well, the marks always confirm my approach, so then I try to figure out what was so good about it.
And yes, sometimes that's very difficult, but sometimes weeks may pass with me thinking well, I really didn't do much. For instance an evasion week. When I get my year planner, I always make sure I have back-up weeks where the learners get a multiple choice task. Well, a week like that doesn't make a lot of sense to me, but sometimes I just can't think of any fun stuff to do and those kids actually find those weeks quite interesting. That's rather odd. And then I suggest: 'Well, shall we do another?' And they reply: 'Hmm, not right now no'. Why don't we do something fun instead'. Something fun, yes.
(3-1-27)

Pete almost wishes to be told precisely what knowledge, skills and understandings are involved in a construct of communicative language proficiency.

I: Yes. .. Well, knowledge, insight and skills. The question is there with a purpose. Because everybody talks about it, don't they? But what exactly is knowledge? What is insight? People might resort to vagueness, but vagueness tends to be disappear when one specifies what exactly one means.

P: Hmm, hmm. Vague in the sense that it's elusive [That's true, that's true] and elusive in the sense that you can't just give it to the children. You've got// Is the tape full?

I: No, oh no, please carry on, I' m just checking it.

P: Yes, yes, yes. If the Education Secretary would hand me a list of words that she thinks the kids should know, well then I'd be done. But that's not the case, and that's vague. I find it vague. All that crap about communicative skills and this and that and then I'm thinking 'Yes, that sounds pretty nice' and then we're all just running in circles. And what does that mean?
(3-1-28)

On the one hand Pete feels that specification of complex knowledge, skills and understandings is complex, on the other he expresses the need for a specification of some kind by the Education Secretary. In chapter 12, we will discuss the need for further specification of complex knowledge, skills and understanding when we review the outcome of this study.

An effective written English language test

When Pete is asked to define the qualities of an effective written English test, he warns that his answer will somewhat 'vaguer' than his response on the other issues he was asked to respond to. Nevertheless, he thoughtfully and cautiously mentions what are essential qualities to him.

P: So now it's becoming a bit vaguer. … I find that it should be communicative. It should deliver a message. … I think it should connect with what you've taught or with what the learner has taught himself. … I prefer for it to … to actually contain something. For instance, when testing grammar, you could do that through very abstract, jumbled, nonsensical sentences, but you could also provide the test with a certain textual content, where the test actually represents something concrete. That the content of the test itself is all right and which makes it more interesting than nothingness.
(3-1-20)

Pete feels that the input as well as the output of effective written tests should be geared to communication, meaningful and valuable to both the learner and the teacher. A second quality Pete mentions is that he thinks effective tests ought to have content validity. As an example he refers to a grammar test. The grammar test we will discuss later in the year actually is *The Little Boy* grammar test, which promises to be about more than mere nothingness.

Interpretations of washback

Pete defines washback first in terms of (de)motivation of marks. Marks can encourage and they can discourage. Even though Pete acknowledges that marks are important for learners and may provide diagnostic information to them as well as to the teacher, he does not see language tests as powerful tools to influence what learners learn and how they go about that. In a way, he sees all the tests as preparations leading up to a kind of final assessment in the years the adolescents graduate. This view is not unlike Mark's in the first interview.

> *I: All right. Now one of the core questions. The role a test plays in learning to communicate in English. In how far do tests help learners learn to communicate more effectively?*
>
> *P: Hmmm. When a learner gets a fair or satisfactory result, a test provides a certain amount of satisfaction. It motivates. It makes them feel content in that 'Gosh, I can actually do that'.*
>
> *I: Self-confidence.*
>
> *P: Yes. Right, that's the word I was looking for. And, at the same time, that test can provide the learner who scored an unsatisfactory mark the insight that 'Maybe I didn't do anything" or 'Shucks, I really worked hard for it, I thought I could do it and now it appears I can't. What should I do now. Maybe I should talk to my teacher about it.' That's where I come in. And at the same time again, I feel that tests should be subordinate to the process of acquiring language skills.*
>
> *I: It's merely a moment of measuring, a tool to, eventually//*
>
> *P: // Yes, but I never elevate my tests to being the ultimate goal of my learners. I do that with the final result, so guys, in two years we want to see at least a six on your final list, done. We shouldn't make a big deal out of it, but then we don't. Because we fiddle around a little here, and we goof around a little there. And then you try to give them some self-confidence and when a test goes completely wrong, you do feel sorry for them. I also feel sorry for myself, because I had the wrong expectations. But especially for them, because they feel like they're stupid. And you never want to make them feel that way//*
>
> *I: //So you do make a distinction. You say that some learners failing the test hadn't done a single thing.*
>
> *P: Yes, but that's not my problem. That's my opinion. I do talk about it to them.* (3-1-32)

An important belief of Pete's seems to be that teachers should do everything they possibly can to prevent tests from taking away the learners' self-confidence and self-esteem. That is one of his reasons for making tests subordinate to the way in which learners learn to communicate. Besides, he feels that all the teaching and testing is basically done with one goal in mind. The ultimate goal is to see at least a 6 on each and every learner's examination score lists.

Pete realises that marks are important for learners. He also acknowledges the potential of both beneficial and negative test washback. A quote with which we will finish our report on Pete's view of washback is an instance in which he deliberately has his learners learn by trial and error. The subject was how to write reviews.

> *It has to say something about the content, but you have to think for yourself what exactly you want to do with that information. The criteria that I use to check it, the lay-out, and what some learners do, you know, they start talking about the ending. I hate it when they do that, but I won't tell them beforehand.*
> *And then they say "Well, you could have told us ahead of time." Yes, but hold it right there, you could have thought about what you were going to do as well and maybe that would have been better." So," Well, well", they say, "so the ending, we shouldn't mention that the next time we do something like this." "No, I don't think you should." And then you always have one learner who goes "And what if we have to write a review" and then I'll say "Right, right, so you would give the ending away, then who's going to read the book? Will that ever happen?" "Oh, right, right, I guess that's true."*
> (3-1-19)

Thus, this particular test may have taught learners who gave the plot away of the movie or novel they were supposed to review that they are supposed to think for themselves and that they are at least partly responsible for the results they achieve. Assessment is never a matter of merely reproducing what a teacher expects of the learners. In the eyes of Pete, it is important for learners to learn to think for themselves.

We will now turn to the final two sections of our first introduction to Pete.

7.5.5 Pete's plans and intentions for the school year

When the first interview was halfway, Pete spontaneously mentioned some expectations of the year to come. The expectations may help us interpret Pete's plan and intentions for the year to come. Just as was the case with some of Mark's statements on basic secondary education, Pete's words would prove to be prophetic.

> P: Well, well, I don't think much will change for me. Well, hardly anything. So I think it won't really make much of a difference as far as English is concerned, but my fear is that the subject will become more difficult for the learners because of the amount of subjects they're getting. There will be more zap-work.
> But I don't know how it will turn out in practice. I don't mean to say that I'm negative about it, or anything, but I just don't know how it's going to turn out. I fear that with all of that planning of time, which is kind of what we're doing right now, that it's just going to be a little too much. And their attention has to be divided over more things and it has been more difficult for them to choose the subjects that got them satisfactory marks in the third form, because they have now opted for profiles instead of for a variety of school subjects. Let's call a spade a spade. Kids don't choose a profile because they want to become a doctor so badly. Yes, we would like them to, but they got a four in Maths in the third form, so it doesn't make a lot of sense to go for a nature and health profile. They mainly choose subjects for which they scored satisfactory marks and in the old situation you could always count on Dutch and English. But now they're getting a second language as well at havo-level. And I'm really worried about my position, not my own, but that of my subject. When you used to score unsatisfactory marks for English, you could still compensate them with others, you know. But it's going to become more important now, because the requirements have been stepped up quite a bit. And my responsibility of making sure that those kids get satisfactory marks, on average at least, has increased. I can't allow myself to make a test which

> is failed by 25% of the learners. Which makes me fear that our standards will start to erode. And in this case I'm not talking about moral standards, but I'm referring to the caesura between what is satisfactory and what is not.
> (3-1-18)

Even though Pete says he is convinced there will hardly be any changes for him personally and for English, he predicts that life is going to be much harder for his adolescent learners. Here are the decisions and plans mentioned by Pete. One of the issues he mentions is the fact that it will become more difficult for learners to compensate for unsatisfactory marks for English. Pete feels it puts more pressure on him to make sure his tests will be passable for most of the learners. Time and data will tell whether this will indeed be the case. We will return to these matters in our discussion chapter.

Pete's plan and decisions for the school year that is about to start are modest. The testing plan has already been laid out. He hopes to visit other schools and colleagues to see how they are coping with the changes to come, but adds that he will probably be short of time. In the end, he returns to one of his worries about the second phase.

I: Your decisions and resolutions for the coming year, Peter?

P: Yes, what do you mean by decisions and resolutions?

I: Well, concerning testing, you know, because you yourself mentioned//

P: //That entire testing plan that we laid out for 4 havo.

I: And that's basically what you've done before, or what you were used to doing?

P: Yes.

I: So no special resolutions because you've officially entered the second phase now?

P: No.

I: Well, you've provided a convincing explanation as to why.

P: That's what I thought.

I: You started with 78 //

P: //I, I …My resolution is to visit some other schools to see how they're managing. But to be honest, I don't believe we're doing such a bad job, if I may say so.

I: I'm repeating what you said, you know, about a good test being a test you yourself believe in the first place//

P: Yes, but this tape can't see the expression on my face, but you know what I mean. That's basically all I'm hoping.

I: Yes, yes.

P: Because the difficult thing about the second phase for me is that I'm being forced to clearly lay out the standards beforehand, clearer than before. And you can do that with listening, you can do it with multiple choice, but I truly believe you can't do that with speaking and writing. In the sense that you're

going to combine two incomparable entities. Namely, a child's intelligence and the content, the communication.
Some people can communicate better than others. But now you have to force this child who can't talk, you have to get him to at least do the basics. And then it's fine, it's good for the child and the person who is like me, who just blurts it out and just does it. I think that person should do more to get that mark, I think, because he should be trying to give the best of him and he shouldn't be going for the basics.
And I truly believe that, when I take writing for instance, I don't believe there's one human being who can check that objectively. And if there is, then I'll bite the dust, but I still don't believe there is. I don't believe in those CITO-tests either. Well, I accept their reality, but I don't yet believe in the objectivity of those CITO writing tests.

I: The example you provided is clear, for sure //

P: // Yes but I haven't seen them in a few years. I still have to ask for them for this year. But then I think they're just too far removed from the learners. I truly believe that those tests are used like, we're teaching and 'Well, right guys, there's also a writing test. Well, here's an example. Why don't you glance over it at home. Just see what you can do with it, but we're going to continue working on other stuff'. And then something entirely different wriggles its way in and next thing you know you've got this test, whereas when I practice it, we'll use a subject like unwanted teen pregnancies, for example. And then over the course of the project they'll be getting more and more information. That information will lead them to process it somewhat into something that works for them, their thoughts and feelings.

I: Which they then express in their writing.

P: And that subsequently go into their writing.
(3-1-23)

Pete believes that it is impossible to assess the productive skills objectively, which is one of the requirements of the second phase. The first interview with Pete finished with a proposition he wants me to react to. As a final question, I had asked my teacher respondents whether all matter they had written down after reading the interview guide had actually been put forward in the course of the interview.

P: I miss one thing. .. A proposition.

I: A proposition (laughter), you're saying?

P: The multiple choice exam doesn't meet the requirements of communicative language education.

I: It did come up during the interview.

P: Yes, yes but not//

I: //but not explicitly. So then//

P: //Do you agree with me or not? You don't?

I: Well, you're asking me a very difficult question. To answer that, I really should turn off the tape.

P: Yes, why don't you.

I: Because, maybe you kind of noticed it a little, but my personal opinions are irrelevant in this investigation. You're starring the show, and there's no place for me on the stage.

P I've been trying to draw you out on what you believe a couple of times, but you never took the bait.
(3-1-33)

Pete could not imagine how difficult it was to maintain the position of a non-participating observer. In chapter 12, we will finally accept Pete's challenge and discuss and comment on washback effects of the CITO reading comprehension national examinations, which are so characteristic of Dutch foreign language teaching and testing.

We will end our report on the data of our first interview with Pete with a reference to the language tests he was to select for discussion in the year to come. The section is remarkably short.

7.5.6 Pete's language tests

The three language tests Pete and I discussed in the course of the school year were identical to the tests Joy had selected for discussion. This was first of all, because Pete fully agreed with Joy's choice for the *Dear Nobody* writing test on teenage pregnancies and the *Little Boy* test on verbal tense and form. Pete only regrets that Joy had offered the *Irish question* reading comprehension test for discussion. When we discussed this test towards the end of the school year, Pete felt this was a bad test. We will report on Pete's response to the three tests more elaborately in chapter 10, entitled *Pete, the project man.*

7.6 Summary

This chapter was the first of the four data chapters of this study. The chapter introduced Joy, Mark and Pete, the three respondent teachers of this investigation. We first offered a general introduction to the three teachers, in which we went into the reasons why the three teachers were asked to participate in the investigation. In addition, we provided information on their careers and personalities as became known in the course of the year of data collection. Finally, we offered some information on the respondent teachers' schools.

Next, the teachers basically had the floor. By way of direct quotes from the interviews we had just before the start of the school year, we learned about the teachers' *formative experiences and core beliefs concerning teaching and testing*. This was followed by the respondents' interpretations of the three constructs central to this investigation. The teachers gave us their views on *learner autonomy, communicative language education*, and *assessment and evaluation*. The sections on assessment and evaluation zoomed in on the *knowledge, skills and understandings* the respondents considered essential for their learners to learn, their definitions and characterisations of an *effective written English language test*, and their perceptions of test *washback* on learning and teaching, when adolescent learners learn to communicate in English. In the two final parts, we were first informed about any *plans or decisions* the teachers had for the school year which was about to start. The last section was a brief introduction to the language tests the teachers were to select in the course of the school year.

The teacher respondents' core beliefs and the interpretations of learner autonomy, communicative language education and assessment and evaluation will be presented in summarised form at the beginning of the three chapters to come. Chapter 8 concentrates on *Joy*, chapter 9 focuses on *Mark*, and chapter 10 calls

attention to how *Pete* looked back on the three language tests he selected for discussion in the year of data collection.

At the end of each of these data chapters will be discussed in how far the teacher's core beliefs and construct definitions are reflected in her/his assessment and evaluation practice. In chapter 11, we will compare, contrast and discuss the teacher construct definitions and compare our findings to the three theoretical chapters of the study.

CHAPTER 8: JOY, THE BUDDING PROFESSIONAL

8.1 Introduction

In chapters 8, 9 and 10, we will report on the follow-up interviews held with each of the teachers in the course of the year of data collection. Each interview was based on a test the teacher had selected for discussion. The follow-up interviews were first of all meant to gather further evidence of the respondents' assessment and evaluation practices. In addition, the follow-up data were used to collect convergent and/or discriminant evidence of the construct definitions and the core beliefs that had been elicited at the beginning of the school year.

The chapter will start with a summary of Joy's core beliefs in teaching and testing and the ways in which she interpreted the three constructs of this investigation.

In the second section we will focus on the question categories that were asked in the follow-up interviews. The full text of the interview guide that was used in the follow-up interviews is presented as appendix II.

In the third section we will briefly introduce the three language tests that Joy selected for discussion. In the three sections that follow, we will discuss the language tests under headings that have been derived from the interview guides. This procedure leads to five categories, i.e. *justification* of the test, the test *conditions,* the *knowledge, skills, and insights* measured by the test, ways in which the learners have been *prepared* to the test, details on the *construction and use* of the test, and finally, details on the ways in which the test was *assessed* and *evaluated*.

We will end our chapter on Joy with a summary, in which we will recapitulate the contents of this chapter.

8.2 Joy's beliefs and construct definitions expressed in the first interview

In this section we will present a brief overview of Joy's core beliefs and her attempts at defining the three central constructs of the investigation. The three tests that were discussed in the year of data collection will be analysed for discriminant and/or convergent evidence of these core beliefs and construct definitions in section 8.7.

As has been illustrated in chapter 6, Joy believes in:

- the advanced learner and/or beginning teacher developing insight in how the foreign language works the moment she actually starts teaching (1-1-27);
- the benefits of exploring literary texts: the main reason why she decided to study English (grammar and fluency she considered her weaker points);
- making how a language works more of a focal point in her teaching and testing practice (1-1-27);
- teachers developing their own materials and tests based on or taken from authentic texts about themes that appeal to the learners. Joy strongly believes in meaningful texts. (1-1-12);
- teachers cooperating closely, e.g. on what to test at what particular time: exchanging and discussing ideas are motivating experiences for the teacher. (1-1-16);
- tests that call for more knowledge and skills than mere short-term reproduction of discrete elements that have been studied by heart. (1-1-2);

- the uselessness of pattern drills and the reproduction of standard sentences in teaching learners how to learn to communicate in English (1-1-31);
- the uselessness of teaching the transformations needed to turn active sentences into passive ones; recognition of passive constructions is enough for her. Energy is better spent on teaching learners how to use proper tenses and verbal forms. (1-1-11);
- the uselessness of rote learning and reproductive testing of idioms and her choice to test idioms in context of authentic written tests or by way of writing assignments. (1-1-18);
- long-term retention of vocabulary, e.g by having the learners set up their own word files and immersing readers in extensive reading of authentic texts that are slightly more difficult than what they are used to handling.(1-1-6);
- integrative tests rather than discrete-point tests (1-1-6);
- the importance of reassuring learners who are likely to panic. (1-1-7);
- the fact that reliably and validly assessing writing tests is problematic (1-1-23);
- the motivating effects of offering the learners some choice, also in areas they are not intrinsically interested in, e.g. literary texts (1-1-34)

Joy's construct definitions

1. **An effective written English language test**

- N+1: level slightly higher than what a learner can be expected to handle
- transfer-oriented: discrete knowledge and skills have to be applied to meaningful and authentic test tasks

2. **Essential knowledge and/or skills in English**

- ability to formulate questions and negations "in reasonably correct ways" (1-1-11)
- ability to put verbs into appropriate tenses and/or forms
- knowledge of the declension patterns of the irregular verbs
- correct word order

As far as insight is concerned, she feels the effects are there, but finds it hard to define what exactly insight is.

3. **CLE**

- Joy struggles to define the construct.
- She does not accept an 'anything goes' approach that allows any verbal or non-verbal means to put a message across. She advocates development of what she has defined as essential knowledge and skills above.
- Authenticity and meaningfulness are two characteristics that pervade Joy's views and beliefs. However, they are not mentioned by her in relation to the construct of communicative language education.

4. **LA**

- Learner autonomy increases if teachers have their learners reflect more explicitly on what they *are doing*, what they *have to do*, what they *want to do* and what *is required in order to achieve this*.
- The process described above is and should be a gradual one, because learners are expected to be told what to do and carry out tasks and activities without a lot of reflection.

- Full learner responsibility is hard to achieve in upper secondary education. The process is restricted by the demands and objectives of secondary education and learner characteristics that are not conducive to fostering learner autonomy.

5. Washback of tests on learners learning how to communicate in English

- Joy feels that tests are not the instruments to help learners learn how to communicate in English. The actual learning takes place in the period preceding the test.
- Joy primarily sees tests as summative assessments, that is as tools to check whether a learner has achieved what (s)he is expected to achieve at the end of a learning process. She does not mention the role of test evaluation.
- Joy does not actively use the test and its assessment criteria to guide learning and determine what the learners learn, how they go about it, and why they do so in the chosen ways.

Before turning to Joy's data, we will first present general information on the interview guides and questions asked in the follow-up interviews.

8.3 The follow-up interview guides

The aim of the follow-up interviews was to collect information on the tests the teacher had selected. As had been the case with the free-attitude interviews, the interviews were semi-structured. Al least five aspects of a test were meant ot be discussed. The five question categories echo the structure of the first free-attitude interviews on the beliefs and construct definitions of the teachers. For the full text of the interview guide that was used for the interviews we refer to appendix II. None of the teachers had any questions after reading the interview guides or felt the need to write down information before the interview started. The questions that were asked related to:

I. ***Justification.*** The teacher justifies why this particular test had been selected for a more detailed discussion.

II. ***The knowledge, skills and/or insights measured by the test.*** The teacher explains what a learner must know and should be able to do to pass the test.

III. ***Construction and use of the test.*** The teacher clarifies details on construction and use, such as the:

i. information on how the test was constructed;
ii. knowledge and skills the teacher needed to construct the test;
iii. ways in which the learners had been prepared to the test;
iv. test conditions;
v. details on the ways the test was scored and graded;
vi. expectations of what learners (should) do with the test after it has been returned.

IV. ***Washback* on *Communicative language Education (CLE).*** The teachers explain in how far the preparation, administration and

discussion of the test stimulate the learners to listen, speak, read and write in English.

V. ***Washback* on *Learner autonomy (LA).*** The teachers explain in how far the preparation, administration and discussion of the test stimulate the learners to learn more independently and responsibly.

In the texts to come, we will present the most relevant teacher data on the five question categories of the follow-up interviews. Similar to our approach in the previous chapter, we will present quotes from the interviews.

8.4 Joy's tests and interviews

Joy selected three tests for discussion. As a first test, all of the three teachers chose the one they felt to correspond best with their views of effective English language teaching and testing. Therefore, the first test Joy offered for discussion was a **writing test** that closed off the so-called *Dear Nobody* project that Pete and herself had developed and taught. The second test was a **grammar test** based on an authentic text called *Little Boy Growing Old before his Time.* The final test Joy and I discussed was a **reading comprehension test** based on an article taken from Time magazine (July, 12, 1999) on the *Irish Question*. All of the tests were teacher-made. The writing and grammar tests were constructed by Pete and Joy and the reading comprehension test was constructed by one of their colleagues and subsequently commented on and revised by Joy and Pete.

8.5 Test 1: The *Dear Nobody* writing test

The *Dear Nobody* writing test illustrates Joy's core beliefs and concerns in English language teaching and testing. The writing test was intended as a summative achievement test on what had been learned in the course of a two-month project on teenage pregnancies. The project was based on Berlie Doherty's novel *Dear Nobody.* Both the novel and its filmed version were part of the project. The four language skills were integrated in a variety of authentic tasks and assignments that invited the learners to build up their personal discourse on the theme. Joy feels that it is important that her learners be offered opportunities to express themselves in English and by doing so develop opinions of their own and discuss dilemmas with others.

The test offered to the learners was in the Dutch language. It has been translated into English for convenience sake. Care has been taken that the original phrasing was retained as well as possible. The original versions of all of the teachers' tests have been included in appendix 6. This is what the *Dear Nobody* test looked like.

Choose and do <u>one</u> of the three assignments presented below:

1. Write a letter to the author of the novel **Dear Nobody**. Include at least the following information:

a. Tell something about the project as it was 'done' at school .
b. Tell something about your opinion on the movie.
c. Tell something about the importance of reflection and discussion of similar themes.
d. Give examples of other subjects.
e. Tell something about the situation in the Netherlands.

2. Identify with Helen. After a couple of years you decide to send Chris a letter. Include at least the following information:

a. Make up a motive for writing the letter. Ask after his present situation (Some 'guesswork'?).
b. Tell something about how life has been treating you (and of course the child) in the time that has passed. (Feel free to use your imagination a little)
c. Once more you look back on that difficult period in your lives. Also include the role your respective parents played. Are you still behind the decisions made at the time?
d. You suggest meeting Chris again. (or suggest the opposite, depending on what you have written down under 'a'. Of course you have your reasons for this.

3. Identify with Chris. After a couple of years you decide to write Chris a letter. Include at least the following information:

a. Make up a motive for writing the letter. Ask how she and your child are doing.
b. Tell something about how life has been treating you in the past couple of years.
c. Once more you look back on that difficult period in your lives. Also include the role your respective parents played. Can you understand better now why Helen acted as she did?
d. You suggest meeting Helen again. (or suggest the opposite, depending on what you have written down under 'a'). Of course you have your reasons for this.

The test above will be discussed in terms of the five question categories of the first follow-up interview we had with Joy.

8.5.1 Justification

Joy had selected the *Dear Nobody* test for four related reasons (2-3-1).

- It was the test at the end of their first project in the school year 1999-2000, and therefore illustrative of the ways in which she and Pete worked with projects.
- Joy felt the tasks and activities the learners carried out in the project provided building blocks that linked up well with the contents of the final test.
- Even though the *Dear Nobody* project integrated the four language skills, the main focus of the tasks and activities was on writing skills. This was true for 4 *havo* as well as 4 *atheneum*.

- The test prepared the 4-havo learners to the school examination writing test they are going to get in week 22 of 1999-2000. The school examination test will also be based on a project and will be laid out similarly to the present test. Her own 4-atheneum learners were prepared for future projects and writing tests to come.

8.5.2 The knowledge, skills and insights measured by the test

Joy has great difficulty in answering my question on what exactly is measured by the *Dear Nobody* test. My question is followed by Joy's sighs and groans. However, she soon regains composure when she starts telling me about the various tasks and activities that preceded the test. Joy cannot separate the test from other classroom activities, including the ones that came before the project itself. She mentions the initial letter she wrote to her learners at the start of the school year with the assignment to write her a similar letter. She also refers to the way the English verb was taught and tested at the beginning of the school year.

> *I: Well, on to perhaps a somewhat more difficult question, the second part. What exactly is measured by this test? [Joy sighs] When you consider what your learners should be able to know and do? [mmm] This might not be such an easy question, but –*
>
> *J: No, and definitely not with a project such as this. [yes] Pfffffff.*
>
> *I: Well, it's not really about what the project measures, I think, but more about what exactly the test measures.*
>
> *J: Yes, but I find it very difficult to separate the two, because the test is such a logical continuation of the project. We started with some grammar exercises, that was at the very beginning of the school year. It was just a review of the rules and doing a lot of different exercises. We started out with separate sentences and then moved on to working more in a context. Furthermore we also started with, well, write your own letter for a change. We wrote an introductory letter, to introduce ourselves and then the learners replied to our letter, which for us was also a way to measure their initial level of English. We provided feedback on the letters, mentioned in class well this is what you are capable of at the moment, but at the end of the year you will be capable of doing so much more when writing letters.*
> *And that encompasses two things that we always assess in writing assignments, which is the technical part, spelling and grammar, word order, those things, and the other part is the content. Are you able to write well, do you have your own style, can you adapt your style to the manner in which the book was written, can you adapt it to the project, things like that. [yes] So on the one hand we measure, well, the technicalities of writing, and on the other hand we measure whether they are able to write quite a long piece in good English.*
>
> (2-1-3)

In assessing writing assignments, Joy distinguishes between aspects of linguistic competence and the actual written discourse a learner has been able to produce. She refers to linguistic competence as the 'technical part'. Joy feels linguistic competence is relatively easy to unravel, and mentions key elements such as spelling, grammar and word order. By 'contents' she refers to the actual written discourse produced by learners on the basis of all that they had learned before or during the project. In line with her core beliefs, she feels the application of linguistic competence to actual spoken or written discourse is far more important than the ability to apply certain structural rules in isolated test sentences.

Despite the fact that the ‘technical part’ is relatively easy for Joy to assess, she does not have a set correction model for the grammatical mistakes and errors she encounters. She calls this a weakness, which she attempts to remedy by exchanging tests with her colleague Pete to check whether they differ in their assessments.

> I*: Have you for instance developed a fixed correction model for the technical part to help you in assessing the mistakes? For example, a spelling mistake is half a mistake and a certain grammatical mistake is a whole mistake or error //*
>
> *J: No, no, no, we don’t do it like that, no.*
>
> *I: How should I look at it then?*
>
> *J: Well, on the one hand you do look at the number of spelling mistakes and grammatical mistakes. And, yes, this really is the weakness of what is being assessed by this test. Yes, you do add them a little bit of course, but a certain spelling mistake stands out more than the other. So I find it difficult to say, well, I count spelling mistakes as half a mistake. I mean, if there’s an extra ‘l’ in ‘wonderful’ or something like that [mm], then well… Things like that. –*
>
> *I: But in principle you do mark those mistakes, so//*
>
> *J: I mark them, I highlight them. I do have my own system you know. I circle spelling mistakes. If the word order is a bit odd or if the English is a bit strange, I underscore the section with a wave-like line. And with the wrong word order, I will mark an arrow to show where the word should go. And – yes, mistakes in the tense, I just cross those out, because they can be very disturbing. And then I review all of that and base the final mark on that. And so mistakes in the tense and the word order, and literal translations from Dutch, yes, those weigh more heavily than spelling mistakes. But I don’t have a real correction model for that, no.*
>
> *I: Well you’re very clear in what you’ve said, that – this is how I operate… [no]*
>
> *J: What we actually do to ensure that our assessments don’t differ too much is that we sometimes exchange a few token tests and mark those for each other. And it usually turns out that we allot almost exactly the same mark. There could be a difference of half a point. Yes, but usually no more than that. So that works well for us.*
>
> (2-1-6)

Joy largely relies on her feelings to determine how disturbing an error or mistake is at a certain stage of the discourse she is reading. If mistakes or errors do not disturb the contents she may simply correct the mistakes that have been made. A criterion she appears to consider as far more important is whether the contents reflect the atmosphere of the novel and film that formed the core of the project. One of Joy’s main targets is to have her learners learn to express their opinions in English and to do so with a certain depth, particularly in 4-atheneum (segment 2-3-17). In general Joy has difficulty in operationalising writing skills in clearly defined terms of required knowledge and (micro)skills. Nevertheless, Joy’s didactic procedures concerning writing skills carefully work towards her ultimate aim of having the learners express their opinions in some depth. After the initial letter writing at the beginning of the year, writing skills are developed further by having the learners concentrate on more formal assignments that move from written descriptions of houses to offering houses for sale.

> *J: Yes, we did do a variety of writing assignments, because afterwards we continued with the letter writing. [mm] Yes, really they were different kinds of short writing assignments. So on the one hand we were doing those grammar exercises just to refresh the mind, to review the basics again. And on the other hand we wrote letters, they had to describe things, for instance they had to take a picture of a house and then describe that house. But, not just provide a description, but also put it up for sale, making the assignment a bit more formal [yes]. So on the one hand they practise informal writing, but on the other hand they also practise more formal writing. [yes] We distinguished between the two.*
> (2-1-4)

Even though Joy finds it hard to indicate what knowledge and/or skills are required to successfully do writing assignments, she carefully builds up the learners' writing skills. However, the focus seems more on the actual products than on the process of writing.

We also discuss what insight is required to successfully do the test. Joy replies that she attempts to stimulate insight by asking the learners reflective questions or having them do assignments that stimulate expression of their own opinions and arguments. An example of such reflective questions is the following. '*Helen is on her way of making an important decision. What do you feel her decision is going to be and why do you think she is going to decide in that way?*' (segment 2-3-27). Another example of an assignment that stimulates insight is to have the learners write reviews of an English documentary of their own choice (segment 2-3-27). Again, the focus seems to be on the product rather than on the writing process.

Summary of the knowledge, skills and insights

a) **knowledge**: the technical side, e.g. spelling, grammar, word order etc. Not too precise. No fixed correction model. Efforts at creating reliability by occasionally having Pete grade letters she had just graded.
b) **skills**: ability to express your opinion in some depth in convincing written discourse that reflects the atmosphere of the novel, film and the theme of the project, i.e. teenage pregnancies.
c) **Insight** stimulated by asking questions that lead to empathy and identification and having the learners do certain tasks, such as writing reviews.

Knowledge, skills and insight are carefully trained from the beginning of the school year onwards: repetition of grammar points, most notably use and usage of the proper verb forms, from spontaneous writing purely focused on meaning to more formalized writing assignments that were graded for grammaticality and adequate use of vocabulary.

The learners were also trained in applying basic self-assessment criteria of writing as presented in two course books they used.

Joy feels that the activities and tasks that were part of the project served as a proper preparation to the test. Her learners were implicitly trained into using the vocabulary required to express themselves in English on the issue of teenage pregnancies in spoken and written forms. Insight was trained by asking reflective questions in the course of the project on teenage pregnancies. A lot of these questions were meant to have the learners develop empathy with the main characters and compare their actions and behaviour with how they would react in similar circumstances.

In the paragraphs below, we will provide some details on the construction and use of the *Dear Nobody* test.

8.5.3 Construction and use of the *Dear Nobody* test

In this section we will report on the categories as they were part of the follow-up interview guides. We will first provide general information on the *construction* of the test. Then we will deal with the alleged *expertise* Joy feels she needs in order to construct the test. This will be followed by details on the ways in which the learners have (been) *prepared* to the test. Finally, there will be a brief section on the test *conditions* and longer sections on *assessment*, that is how the test has been scored or graded, and on *evaluation*, primarily directed at what learners are expected to do with the test after it has been returned. For the sake of conciseness, we have left out direct quotes most of the time. Instead, we will refer to the relevant interview segments in which the information was found after each paragraph heading.

Construction (2-1-28/29)

The *Dear Nobody* test was constructed by Joy's colleague Pete. Joy fully agreed with Pete's proposal and was particularly pleased with the choice the learners had between three possible letters to write: one rather technical letter to the author listing facts and details related to the project and two letters in which the learners had to identify with one of the two main characters. In her 4-atheneum class, only one or two learners opted for the letter to the author (assignment 1). The boys predominantly identified with the male main character *Chris* and the girls with *Helen*. What pleasantly surprised her was that the boys appeared to be just as capable of empathising and identifying with the main characters as the girls. This was not what she had initially expected.

Expertise *(2-1-29/30)*

Pete had constructed the test and sent a first version to Joy for her to comment upon. This means that the test required no specific expertise on Joy's part. As indicated above, she was impressed by the test and appreciated the version as it was. The test was fully endorsed, because it was in line with her and Pete's core beliefs. It was based on a meaningful project that integrated several language skills. The test invited the learners to express their knowledge of the novel and their views on teenage pregnancies in meaningful ways.

Joy's way of assessing writing was copied from the way in which her university writing had been assessed when she was educated as a professional translator. There, too, distinctions had been made between the mechanics of writing and the content. Any errors or mistakes were represented in a clear and useful way that so much appealed to Joy that she decided to copy it in her assessment practice (2-1-30).

Conditions (2-1-31/32)

The test was done in two fifty-minute classes. In the first lesson the learners concentrated on a brainstorm, using techniques such as a bubble diagram and on writing a first version of the letter. Joy briefly had a look at the first versions of learners who wished for feedback. The second fifty-minute period was spent on the learners writing their final versions. The actual brainstorm was there because it always had been part of the writing assignment the learners had done in preparation for the test. The individual brainstorms were not assessed by Joy. She simply wanted her learners to think about the structure and contents before they started writing.

The use of English-English and Dutch-English dictionaries had been trained for a brief period of time. During the actual test the learners were not allowed to use dictionaries, but Joy adds that the learners will be allowed to use them in tests to come.

The intended length of the letters was in between 400 and 450 words.

Preparation (2-3-5/17)

When we had our second interview with Joy, she handed over to me the materials that had been given to the learners. It was a document that consisted of 5 pages, full of varied tasks and assignments geared at introducing the learners to the novel and theme and develop their reading, speaking and writing skills on the way.

The materials the learners got consisted of an a4 assignment sheet in English and four pages of text. In a two-week period, the learners carried out drama, reading, speaking, and writing tasks based on six authentic sources. These sources were Berlie Doherty's novel *Dear nobody*, the corresponding movie, the film *The Snapper*, a dramatic script taken from that movie, and two background reading comprehension texts entitled: *Teenage pregnancy, the crisis that isn't* and *Teen births make Britain the youth club of Europe*. We will present three sets of tasks from the original eight to illustrate how Joy's learners were prepared for the test. The first task read as follows:

> A. A fragment from: the Snapper.
> Read the extract carefully. In small groups, try to decide on how you would 'act' the different 'parts' (how old, for instance, do you think the twins are?).
>
> In groups of 7, rehearse a reading of the fragment, remembering the things you came up with before.

The extract concerned a dramatic scene in which 20-year-old Sharon Rabitte informs her Irish parents that she is pregnant. Sharon refuses to identify the father and also refuses an abortion. The play was first prepared for by a small group, and subsequently acted out. After that the class watched the original film scene on video. The following two sets of tasks seem to have been particularly relevant to the ultimate *Dear Nobody* writing test at the end of the two-week period:

> B. Watch the first episode of ***Dear nobody***. Then discuss the following points:
> 1. Was Helen wrong to keep Chris waiting for so long without an explanation?
> 2. What do you think of Chris's reaction to the news?
> 3. Why did neither Helen nor Chris tell their parents?
> 4. What courses of action are open to Helen and Chris? Which do you feel is the right one?
> 5. At the very end of this episode, Helen makes up her mind to do something. What do you think she is going to do?

After the introduction the theme of teenage pregnancies, the focus is now on the filmed version of *Dear Nobody.* By way of five questions, the learners are challenged to find answers to the moral questions raised above. The answers are first given individually, then discussed in small groups and finally discussed with the teacher with the whole class. The movie scene is followed by an extract from the novel and two identification questions.

As a final illustration of the preparatory tasks the learners received, we present a note-taking assignment. The focus is again on ethics and morality.

> G. Watch episode three. Then make your **personal** notes about the following discussion points:
> 1. Do you think Helen's family adjusted well to Helen's decision?
> 2. Does Chris have a responsibility as a father now that Helen has broken off their relationship?

3. Why would Helen suddenly decide to send all her *Dear Nobody* letters to Chris?

All of the materials and tasks were meant to help the learners prepare for the final writing test, which was to come some two weeks after the introduction of the project. Joy says the subject of teenage pregnancy appeared to be relevant to the learners from the start.

Assessment (2-1-8 and 2-1-25/33)

Joy marked the test on the ten-point scale commonly used in Dutch education, with 10 being the maximum score. Joy and Pete had agreed on a so-called bottom score for this test. It was impossible for learners to score below 5. The final mark for the test was composed of two marks. A mark was given for the mechanics of writing, such as spelling and essential grammar points. Another mark was given for content. The latter mark was dependent on how well the contents of the letter reflected the atmosphere of the novel and movie. The marks for the mechanics of writing allegedly were considerably lower than the marks for its actual contents.

In deciding on the exact mark to give for each category, Joy in part assesses intuitively and holistically. She finds it hard to assess writing skills. When she assesses the mechanics of writing, there may be some grammatical errors or mistakes she considers as minor. In those cases she simply writes down the correct English on the test paper. Errors or mistakes related to the contents she adds to the test papers as brief written comments. Such errors may refer to abrupt beginnings, lack of clarity about what the writer means to say, lack of logical connections between the various sections in the letter, use of the personal pronouns ‘it’ and ‘ they’ without clarity about what or who the pronouns refer to. She also objects to superficiality and to writing that does not reflect the general atmosphere of the novel or movie (2-1-8).

Joy often starts correcting tests by pencil, giving her the opportunity to change her initial assessment when the need arises, e.g. after having checked a particular assessment with her colleague Pete. Her consultations with Pete and the occasional exchanges of her tests with his seem to reflect her genuine concern with ensuring acceptable levels of inter-scorer reliability.

The way in which she marks tests is by using a marking procedure she became familiar with when she was a student in English and Italian translation. She felt the system was fair and her university lecturers were always able to convince her what kind of error or mistake had been made (2-3-30). Thus, a previous experience with assessment and evaluation that was considered as useful and fair was copied by Joy.

Each of the three writing assignments the learners are able to choose from, consist of 5 (task 1) or 4 (tasks 2 & 3) elements that ought to be part of the letter. If a part is missing, Joy writes some commentary on the test paper and subtracts marks from the final score. She was not precise about how many points would be subtracted for each and every part.

The most important assessment criterion for her was whether a learner’s response fitted the atmosphere of the novel or film and the personalities of the two main characters. One learner had fantasised that Chris had become addicted to drink and wanted to commit suicide. Joy did not accept this, as it was not in line with the way Chris and Helen had been

portrayed in the course of the project (2-1-33). It would be impossible for this person to reach the maximum score. However, a good score would still be in the learner’s reach if his English was fine. It seems to show that Joy prefers appropriateness of language use and the ability to make logical deductions to grammatical correctness, unless the structural errors interfere with the intended meaning.

A part that Joy felt was rather difficult for her learners to write about was the first part, in which they had to make up a motive for writing the letter after such a long period of time. Havo learners generally experienced more difficulty here than her 4-atheneum learners. All of the other questions did not cause any specific problems and were relatively easy to do.

Despite the intuitive and holistic nature of her assessment, Joy is comforted by the fact that in retrospect she is always able to explain to learners how she arrived at a certain score. She claims none of her learners has ever challenged the fairness of her assessments. There are three likely interpretations of the fact that her assessments are generally accepted. These interpretations are based on our classroom observations and informal conversations we had with some of the learners. The learners seem to willingly accept Joy's judgment, partly because the learners have become familiar with her way of assessing writing skills in previous writing assignments, partly because the learners had learned she could be trusted, and finally because the test was seen as a summative event and the marks were generally okay. The learners apparently felt no need to concentrate on the specific knowledge and skills anymore. New projects would create new, and perhaps different, opportunities.

As a final remark on assessment, Joy mentioned there were no significant differences between the results of her 4-havo learners and her 4-atheneum learners on this test. Both level groups did just as well and were assessed in the same way (2-1-25).

So far, we have concentrated on Joy's assessment practice as it was illustrated by the *Dear Nobody* test. Because we specifically distinguished between assessment and evaluation in chapter 5, we also analysed the data for evidence of evaluation practices.

Evaluation (2-1-34/36)

The evaluation details that Joy offers primarily relate to the activities and tasks that preceded the *Dear Nobody* test and a single test in which the learners' writing skills had been tested before. In the case of the *Dear Nobody* test, neither Joy nor the learners appear to have felt the need to closely evaluate the scores and results of the actual test. What follows below is an account of the ways in which feedback is given to the learners to have them develop their writing skills.

In doing assignments, Joy's learners largely work on their own. All of the assignments and tests they do in the course of their training have to be put into a file readily available in the classroom for both the learner and the teacher. It enables Joy to provide formative feedback in the course of their training, if she can find the time. Unfortunately, this time is often lacking. She only checks the files occasionally, and is then able to put a few learners on the right track again.

Joy had undertaken two efforts to stimulate peer assessment and evaluation. First, she had her learners study information on how letters should be structured and built up in *Touch down,* the course book that was used for a short period of time. Later a checklist (Mulder, 1997:113) was copied for the learners to assess their own and their peers' writing. Joy states that the learners had great difficulty in applying the listed criteria. She mentions two problems. First, not all of the groups put a lot of effort in trying to assess their own writing and the writings of their group members. Second, a lot of learners felt insecure in using the list and regularly asked for her feedback (2-1-9).

8.5.4 Washback on CLE

At first Joy found it hard to answer my question on how the test affected how learners learn to communicate in English. When I rephrase my question she mentions fluency training and the training of writing skills as two essential components of the project.

I: The question seems more difficult than it really is, because you have of course already mentioned quite a few components of the project.

J: The link with, yes, I just don't know//.

I: Let me return to the interview, you know, the one about communicative language teaching. [yes] We// Our task is to teach learners how to communicate [yes], in English. What are distinct , communicative ingredients for the learners in this project?

J: Yes, I think, I think, all aspects of it. [Yes, I agree]
I think especially the speaking component, talking to each other about what they saw, what they, also about what they read.

I: They did that in English as well? [yes, yes] And they spoke about it in English?

J: Yes. [yes] And yes, the component of thematic writing. [yes] You write about it, you identify with something or someone or you write about the project or something.

I: So that actually amounts to two moments of production, after the input of English, right? [yes] That was also a part of the project. [yes] Did all the groups manage to actually speak English?

J: No, no, not all the groups. Yes, (laughter) we did always try to. [okay] We also tried to do the entire project in English only [mm], but that, for me, for me that remains my weak point (laughter).

I: To do everything in English?

J: Yes, to do it in English all the time, because there are always learners who revert to Dutch and I go along with that, I have noticed. [mm] Very easily. [mm]
(2-1-37)

The segment above seems to indicate once more how difficult it is for Joy to mention concrete knowledge and skills that make up communicative competence. Although the *Dear Nobody* project was full of interesting communicative tasks, washback of these tasks on the ways learners learn how to communicate in English was not on her mind.

Joy does not provide a lot of detail on the components in the project that stimulated communication. Her response echoes the difficulties she experienced in the first free-attitude interview, when Joy was asked to define communicative language education. She is somewhat more specific when it comes down to washback on LA. This is how she responded.

8.5.5 Washback on LA

Joy's efforts at creating positive washback of the *Dear Nobody test* are that she tells the learners that they have to consider the mistakes and errors that have been made and to correct the ones that have not yet been corrected by her. It appears that some learners do this, others do not. Joy considers this the responsibility of the learners themselves. They can always come and see her when they feel like. Joy is not going to check whether the learners have actually corrected any errors or mistakes, let alone have learned from them in view of tasks and activities to come (2-1-35).

A part of her teaching that was not specifically mentioned by Joy as having a washback effect was that the *Dear Nobody* test was preceded by a writing test that had been scored and graded in ways similar to the *Dear Nobody* test. It was to familiarise the learners with the test format and ways of scoring and grading, which may have had a positive effect.

The interview on the *Dear Nobody* test provided additional information on how Joy saw LA.

- gradual move from direct teacher instruction and individual and group classroom tasks closely monitored by the teacher to more freedom and independence for the learner.
  Joy regrets the freedom given to her learners in a second project. As a teacher she had completely lost track of where the majority of her learners were at before the knowledge and skills of this second project were tested.
- Learners are supposed to be helped in their planning by giving them so-called study guides. The guides focus on *what* has to be studied or done at *what* particular time. Joy does not think that teaching learners how to plan is the task of the subject teachers. She feels this is one of the responsibilities of the mentor teacher of a particular class. The guides that were used in the year of data collection did hardly ever focus on the *how* and *why* of the tasks and activities that were meant to be carried out.
- Joy feels that learning on your own involves peer learning.

The *Dear Nobody* project was characterised by Joy's direct instruction, which was followed by assignments and tasks the learners were asked to do alone or in small groups. Joy supervised her learners in the classroom most of the time. She refers to herself as a 'control freak'. Particularly at the beginning of the school year she wants to be sure that all of her learners can cope with the responsibility to work and learn on their own.

At the time we had the interview on the *Dear Nobody* test, Joy had just prepared her class for a writing test on the novel *Cal* by Bernard MacLaverty. In preparing for the so-called *Cal*-test, her 4-atheneum learners were given far more freedom in doing the tasks and activities related to this project. Apart from a few learners who had planned to see her regularly, she had lost track of what the majority of her learners had been doing, let alone of what exactly they had learned by doing the assignments. Joy feels less comfortable with this approach than with the way in which she had prepared her learners for the *Dear Nobody* test.

8.6 Test 2: The *Little Boy* grammar test

As a second test Joy chose for a grammar test in which the learners had to fill in the proper tense and form of a given infinitive in an authentic text. By way of introduction to the *Little Boy* grammar test, we will briefly go into the grammatical issue the test focuses on and explain why English tense and form are generally difficult for Dutch learners.

Most learners of English have difficulties when they apply the rule that time adverbials, such as *tomorrow, yesterday, when I was young, …since 1968*, require tenses that are often different from the tenses that are common to the Dutch language. These differences often result in negative interference between the Dutch verbal system and its English equivalent. Below are five examples that illustrate difficulties Dutch learners generally experience in this grammatical area. The first sentence contains an error or errors a Dutch learner is likely to make. The second

phrase is a correct Dutch phrase that was directly translated into English. Finally, there is the correct English sentence.

> I go tomorrow to Paris * -- Ik ga morgen naar Parijs -- I am going to Paris tomorrow
>
> When are you born?* – Wanneer ben jij geboren? – When were you born?
>
> I have lived in Amsterdam for years.* – Ik heb jaren in Amsterdam gewoond – I lived in Amsterdam for years.
>
> I already live in Nijmegen for ten years now* -- Ik woon nu alweer tien jaar in Nijmegen – I have lived (have been living) in Nijmegen for ten years now.
>
> It rains – Het regent.--It is raining.

As a further introduction to the *Little Boy* test, we will compare the actual test with the original text on which the test was based.

We managed to retrieve the original text from the Internet. The article had been taken from the *Daily Mail* of Tuesday, October 13, 1998. The original text had the following tenses and forms in the 29 gaps that had been made:

1. has started
2. will be
3. is suffering
4. carry
5. have refused
6. have raised
7. was diagnosed
8. affects
9. varies
10. gave
11. he has defied
12. are determined
13. has forgotten
14. learned
15. is starting
16. have
17. has shown
18. thought
19. have (we) done
20. do not have
21. handles
22. could
23. does not
24. has gone
25. is running
26. is (already) starting
27. try
28. are looking
29. will not happen

As compared to the original text, only one time adverbial has been added. It concerned the sentence of item 24. This item was indeed a tricky one. It would have been hard to solve for most of the learners if the addition of the time adverbial ‘Up till now’ had not been made. Below you will find the original sentence, followed by the line included in the test.

> Every year Elliott <u>has gone</u> to the National Institute of Health in Washington.
>
> Up till now Elliott **…24…** to the National Institute of Health in Washington every year.

The addition of the time adverbial “Up till now” is to help the learners use the present perfect tense of the verb “go”. The time adverbial of the original text “Every year” would have lured a lot of learners into using a simple present tense.

The test presented to the learners by Pete and Joy contains some irregularities. First there were two shortcomings that have probably not affected the learner score in any serious way. The line that contained item 8 ended in a comma, instead of in a full stop:

> Niemann-Pick usually **…8…** school-age children, but can strike at any time,
> Life expectancy normally **…9…** between five and 40 years.

A second minor mistake was that item 29 was listed twice: “29. Look” should have been “28. Look”. Because the text did list gap 28, we suppose this mistake did not do much harm either.

Some of the other irregularities may have been trickier for the learners and could well have affected their self-confidence, intuitions or scores. Item 7 is presented as “Be diagnosed”, a passive which includes the auxiliary “to be”. This is in line with Joy’s and Pete’s beliefs. Both teachers believe that passive constructions only need to be recognized by the learners. Active use of the passive is not required. By presenting the passive as “Be diagnosed” instead of the option “diagnose”, which would have had our preference, the learners are helped because the actual passive construction is already half there and its auxiliary given. So far, so good. But a little to my surprise, item 12 lists “Be determined”. The suggestion here strongly is that this is also a passive construction. However, “Determined” is an adjective preceded by the copula “to be”. The construction “We are determined” resembles a sentence like “The door is open”, whereas “The door is opened at eight every day by the manager” would have been a passive construction.

Another error is item 11. In the phrase ‘, but **…11…** them so far.’ the subject ‘he’ is missing. This may well have confused some of the learners, particularly those who do not the meaning of the verb ‘to defy’.

Another irregularity concerns item 19. In order to fit in well with the text, the infinitive ‘do’ has to be transformed into ‘have we done’. Because the subject “we” has already been given, we expect that most testees have filled in ‘have done’.

> “Why us, what we **…19…** to deserve this?”

A final item that invites commentary is number 22. The auxiliary ‘Can’ is presented. It is the only modal auxiliary used in this test. This is a complicated way for soliciting the use of the modal “could” here. Our guess is that some of the learners have simply copied “can”.

Despite these minor confusions, which are not uncommon to teacher-made grammar tests, the test is an interesting and admirable effort to assess grammatical tense and form by way of authentic texts.

In the sections that follow, we will again use the structure of the interview guides.

8.6.1 Justification

Joy selected the *Little Boy* test for four reasons.

- For lack of an interesting alternative test. The only likely alternative for the grammar test would have been another writing test, similar to the *Dear nobody* test she had discussed before.
- As compared to the *Dear Nobody* test, Joy feels the grammar test measures more adequately the relationship between test preparation and test performance.
- As this test was a discrete point test, it was easier to assess in a straightforward way which response was right and which was wrong.
- It was interesting for her to assess how her learners were able to apply rules related to the English verb to an authentic text.

Joy considers the last reason as her most important justification for choosing this particular test. We have already learned that she does not believe in simple reproduction of grammatical rules in isolated test sentences that often have no relation with real communication. A grammar test based on an authentic test enabled

her to assess whether learners are able to apply rules related to a heuristic scheme of the English verb she had dealt with at the beginning of the school year.

8.6.2 The knowledge, skills, and insights measured by the test

Again, Joy is slightly puzzled at first when asked what exactly is measured by this test.

> *J: Erm, yes, erm. What it measures? I think that this test measures whether they can use their grammar, whether they can apply it to English as it is used, say, in, in newspaper articles.*
>
> *I: And with that you mean// What kind of grammar do you mean?*
>
> *J: This kind, look, I believe that this article is, it isn't very formal. I think, I believe it is pretty communicative English. In the sense that you would also hear the language being spoken like this if you had a conversation with someone.*
>
> *I: Do you remember where this article is from or is that//*
>
> *J: No, I don't really. I think it's from a magazine.*
>
> *I: How interesting that you say that the kind of grammar also depends on the kind of English that you are using.*
>
> (3-1-15)

Joy's belief 'that you would also hear the language being spoken like this if you had a conversation with someone' slightly puzzles us. Even though the text is accessible and attractive, we would not have labelled the English as typically everyday. Joy then explains that she feels that authentic English should preferably include informal everyday English. We then return to the knowledge and skills that the test intends to assess.

> *I: All right, fine. Yes, the kind of grammar that you test does that in fact focus solely on verbs or do you assess more than knowledge of verbs?*
>
> *J: No, it does focus on the verb. I mean, adverbs and things like that aren't tested in this case. [OK, yes] It's solely, yes, no, it's solely the verbs.*
>
> *I: What caused you to choose for verbs?*
>
> *J: Because I think it's more disturbing when verbs aren't used correctly in English than when you occasionally forget to use an adverb, for example. Testing word order is also important, by the way. It's not explicit in this test, but it did come up.*
>
> (3-1-15)

Joy says that the test assesses whether the learners can apply their knowledge of the English verbal system to a number of given infinitives taken and selected from an authentic text of an English newspaper or magazine. Both the time adverbials and the context of the sentences provide the clues the learners need to apply the so-called 'when scheme' to the given infinitives. This is a heuristic scheme that helps learners to recognise time adverbials and subsequently opt for the appropriate tense and/or form. Joy insists that the meaning of the text is just as or even more essential than the scheme that has been studied. Unfortunately, she was not able to indicate concrete examples of items that required context and meaning rather than the application of the verb scheme to arrive at the correct responses. Joy again states that the text itself should not be too formal or highbrow. The English should be the English you hear in the streets.

The English verb was not the only aspect of English grammar she had concentrated on at the beginning of the school year. She had also dealt with word order and matters such as the differences between adjectives and adverbs and recognition of passive constructions in English texts. Yet, she had chosen for testing English verb forms because she considered mistakes or errors in this area as disturbing.

Joy feels that knowledge of the text is required in addition to knowledge of the 'when scheme', tenses and forms, and the irregular verbs.

> *J: Yes, you have to make sure you understand a text. Otherwise you can't do a test like this at all. Knowledge of the rules is not enough. You have to look at the context as well.*
>
> *I: How important is an understanding of the text to the learners, would you say?*
>
> *J: Well, it is very important indeed. Because sometimes there is a clue that helps the learners select the appropriate tense or form. This clue may be somewhere in a paragraph, and not in the actual sentence itself.*
>
> *I: So learners really have to read the full text?*
>
> *J: Yes, you have to read on and understand what is being said.*
>
> *I: So you think it is impossible for a learner to do well on the test by only using the 'when scheme'?*
>
> *J: No, it doesn't work like that. It was clearly shown that it doesn't work like that. It is not a matter of learning a scheme or certain rules by heart. You really have to look further than that.*
>
> (3-1-16)

Summary of knowledge, skills and insights

Knowledge: The knowledge that Joy feels is required to do well on the test is knowledge of:

- English tenses and verbal forms;
- the verb scheme: a heuristic 'when scheme' offered to the learners;
- English irregular verbs.

Because Joy feels that the context of the text provides additional clues about using the proper tenses and forms, the learners should be able to understand the text. This means that they should master:

- the vocabulary and meaning of the text;
- understand the ways in which sentences and paragraphs are linked together.

Joy feels that proper and authentic language use invariably means that a variety of knowledge and skills are integrated in some form or other when classroom knowledge is applied to the real world (3-1-28). In that sense, the *Little Boy* test can again be referred to as an integrative test.

Skills:

- the ability to find clues, such as adverbials, that tell one which particular tense or verbal form has to be used;
- reading comprehension (as defined under 'knowledge' above).

Insight: Joy feels that insight is measured indirectly in the test. She feels that the reasoning process during test preparation may have helped some of the learners to become more insightful regarding the application of the when-scheme (3-1-25).

Joy's reference to the reasoning process is a reference to another skill as well. The ability to reason or justify why a particular tense or form has been used had been part of test preparation over and over again. Despite the attention it had got during practice, process questions had not been included in the actual test.

In the next section, we will again present details on the construction and use of the test.

8.6.3 Construction and use of the *Little Boy* test

Construction (3-1-27/32)

Joy and Pete have been testing verbal tense and form by way of authentic gap-fill texts for two years now. Before that time, they wrote their own gap-fill texts. In search for more meaningful contexts, they decided to use texts from magazines or tabloids that were accessible to the learners and not too highbrow.

Joy discovered that using authentic texts was an attractive alternative to writing their own texts. In line with one of her core beliefs, Joy prefers this type of test to grammar tests that are based on isolated sentences. She feels rules of essential grammar, such as verbal tense and form or word order, have to applied in meaningful contexts. Although she was already partial to the idea of integrative grammar testing, it was Pete who introduced her to this type. The grammar tests were subsequently constructed in partnership with Pete. If required, minor adaptations would be made.

Expertise (3-1-33/39)

Joy first explains that what is primarily needed is the right frame of mind. She refers again to two of her formative experiences. Joy had already experienced the uselessness of focussing on mere reproduction of knowledge and skills when she was at secondary school. In addition, she had really started to reflect on how the language works when she was at university. Analyses of her own written and oral work with others, had made Joy conscious of the role of grammar in oral and written production. Lectures on e.g. syntax had not created that effect. This had again been a matter of reproducing what had been learned. Her university experiences coincided with her experiences when she had to explain English grammar to a relative of hers. However, this experience was not mentioned in this interview.

So the willingness to test grammar differently from traditional discrete-point tests had been there from the start of her career. She had made a modest start with alternative types of testing by writing her own dialogues when she first started teaching. She had started thinking about these matters when she was at university.

Joy feels that assessment procedures such as the *Little Boy* test are not too easy to construct. A first issue is to select texts that are not too difficult. She feels this is particularly important for fourth-formers. The selection of such texts is a process of trial and error. Gradually, she has developed feeling for selecting appropriate texts. She feels the first texts offered to learners should be relatively easy and preferably taken from magazines. A second reason why this type of test is hard to construct is that the texts need to be adapted here and there. Sometimes, time adverbials will be added. The alterations may also concern some idioms of an authentic text.

Conditions (3-1-41)

The test was administered in a 50-minute session. Dictionaries were not allowed. Joy adds that dictionaries had not been used in classroom test preparation either.

Preparation (3-1-2/4; 3-1-40)

The learners have been prepared to the test gradually and intensively by way of assignments that increasingly required peer cooperation. First, Joy explained the 'when-scheme', a heuristic scheme that helps the learners find time adverbials or the any implied references of time and consequently select the appropriate tense and form of the verb. Then, she starts in the traditional way the learners have been used to, by offering them discrete-point translation sentences with easily discernible time adverbials, such as 'Ik woon al tien jaar in Deventer' (I have lived/have been living in Deventer for ten years (now)). After that she had her learners study grammar rules and irregular verbs from the course materials they had been using at the beginning of the school year. The learners were then asked to construct translation sentences for one another and discuss the translations in small peer groups. In the meantime, Joy was walking around to help out when necessary. When the learners worked in groups it was important that they were always able to explain to one another why a particular form or tense had to be used. After that the move was made to practising with authentic texts taken from magazines. The verb forms had been omitted and replaced by infinitives. The learners had to reason what tense or form the infinitives should be put in. The texts gradually became more difficult and complex.

The key elements of test preparation have been determining the learners' initial situations, direct teaching of the 'when scheme', and guided practice followed by peer practice, all with materials of increasing authenticity and complexity.

Assessment (3-1-42/48)

Joy had worked on a key with the original forms and tenses and with some acceptable alternatives before the test was administered together with Pete. Having given the test before, they also agreed on a norm of 2 errors per point deducted from the maximum score of 10. Joy came across acceptable alternatives when she was marking the test. The alternatives were discussed with Pete, and generally accepted as correct answers. The norm was similar to the norm Joy and Pete had used the year before, when the same test was administered as well. The learners had been familiar with a norm of around 2 errors per point in their practice with sample tests.

Joy says that the items that were generally most difficult were the ones that did not have explicit time reference.

Evaluation (3-1-49)

Joy tells me that she returned the tests and first had their learners correct the errors they had made as a homework assignment. In the next lesson, Joy reviewed the test by basically giving the correct response to the items and explaining why the verb forms and tenses were the correct ones. and giving the learners who had done their homework assignment the opportunity to ask questions.

I: You have returned the Little Boy test. Have you reviewed the test with your class in any//

J: Yes, I think I had them correct all of their errors at home. After that we reviewed the test at school.

I: And did you discuss the test in full? Did you review all of the items?

J: Yes, I did.

I: Was their any specific way in which you reviewed the test?

J: Well, yes, I simply gave them the correct options for each of the items. The learners had already thought about the correct response at home. I myself then told them why a particular verb or tense was the correct one to fill in. The learners were given the opportunity to react if they failed to understand me
(3-1-49).

In Joy's way of reviewing the test, the responsibility is basically with the individual learner. The learners decide whether and/or how they are going to review the test. It is up to them to ask questions if matters are still unclear.

We will now turn to Joy's perception of any washback of the test on communicative language education and/or on the nature and degree of autonomy of the learners.

8.6.4 Washback on CLE

Joy feels the positive washback effects of this test on the learners' communicative skills are practically negligible.

I: In how far does the preparation, administration and review of this test stimulate how the learners learn to communicate in English? … So in how far does it affect their listening, speaking, reading, writing, or whatever they need to be able to do in English?

J: Well…

I: It isn't a simple question, is it?

J: Well, no. It may stimulate it a little, but I feel the effect is negligible.

I: That means that you would opt for other ways [Yes] or different content to attain communicative goals.

D: Yes, I think I would. Like I've just said, I think that extensive reading is far more important than the effect of testing grammar rules. Yet, it is useful to know how things work….I feel
(3-1-52).

Joy mentions a future focus on extensive reading, listening, and on vocabulary training in view of the disappointing results of their present fourth formers on regular listening and reading tests. The fourth formers this year had scored far below the scores that fourth formers usually get.

Nevertheless, Joy will stick to a grammatical focus at the beginning of the school year. Earlier on in the interview, Joy had mentioned a more or less positive washback effect on motivated learners who really wished to prevent making grammatical errors in the English verb in the writing examination to come. For Joy herself, there is a positive washback effect. In the case of future verbal errors, she only needs to refer to the when-scheme and all of the sample sentences and texts the learners had done at the beginning of the school year. However, she realises that only a small number of motivated learners will actually look again at the practice sentences and texts as part of their preparation for the writing test.

J: (After the first six weeks) we haven't tested grammar anymore, that is not so explicitly. Yet, we did assess their knowledge of verb and tense implicitly, in the case of writing assignments and the like. In those cases we referred again and again to the 'when scheme' and any persistent errors they had made at the beginning of the school year. We encouraged them to look back

on their materials if they felt uncertain about their achievements. But we haven't tested grammatical knowledge explicitly anymore.

I: You haven't, OK. and this way of working is comparable to what you are used to doing in the fourth forms? That you focus on a recap of grammar at the beginning of the school year?

J: Yes, exactly. And we'll be doing that again next year.

I: I see. So you really appreciate this way of working.

J: Yes, also because the learners will have a writing test as an official school examination at the end of the year. And, yes, this has been integrated in a project. So learners primarily concentrate on the content. But if they need to concentrate on the technical side and focus on some weak grammatical points, we can again refer to what has already been dealt with in class. Take your Touchdown handouts and study the grammar notes once more. Look again at the 'when-scheme', and, most importantly, look at the exercises and assignment you did at the time. Where and when do you make errors? Take any previous writing in consideration, in which you will see that you will make the same kind of errors again and again. ….Probably

(3-1-10)

Joy says that it is not so much the actual test, but all of the notes, exercises, and assignments that were used in test preparation that may help the motivated learner to look back at materials once more. Whether the learners actually follow up her advice is left to their discretion.

8.6.5 Washback on LA

Joy feels that the ways in which the learners had been prepared to the test called for little learner autonomy. All the exercises, assignments and texts were carefully structured and staged. The when-scheme was given to them and explained by the teacher. After that numerous exercises followed in which learners had to explain to each other again and again *why* a particular form and or tense was the correct one. Joy feels the 'why question' is an essential component of the when-scheme. Process-oriented questions had been part of test preparation, but did not play a role in the actual test. The *Little Boy* test was product-oriented. Joy did not warm up to my suggestion to include supplying the reasons why a particular tense or form has been used by for three or four of the thirty items tested.

I: When you prepared learners to the test you distinguished between the actual response and the reasons why they opted for a particular verb or form, because you felt this was important. Yet, you haven't decided to ask, for example some three of four times, for reasons why a particular verb tense or form had been chosen.

D: No, no, no, no. There was room for that in the phase that preceded the test, and then ultimately erm .

O: It should really be part of test preparation? [Yes, I think so] and the test is the final product.

D: Yes, really the product, yes.

(3-1-26)

Joy feels that the effects of the *Little Boy* test on the learners' communicative abilities and on fostering their autonomy as learners are very limited. However, she does feel that the attention paid to verbal tense and form is beneficial to the

motivated learner in the course of the school year and is beneficial to her as a teacher, because she can refer to a period of about six weeks at the beginning of the school year, in which the focus generally is on tense, form and their occurrences in authentic written English.

As a third test selected for discussion, Joy came up with a reading comprehension test. Not surprisingly, the informal test was again related to a project that Joy and Pete had developed and taught.

8.7 Test 3: The *Irish Question* reading comprehension test

As a final illustration of her assessment and evaluation practice, Joy opted for a discussion of a reading comprehension test. In chapter 6 we already explained that reading comprehension is the skill that is tested by way of a national examination for the foreign languages in the Netherlands. The graduation score for English is made up of the average score of the school examination and the score on the national reading comprehension test. The national test consists of a mix of English multiple-choice questions and Dutch questions on reading comprehension that have to be responded to in a concise and clear way. That is the reason why the assignments to the *Irish Question* reading test are in Dutch.

The test consisted of two sheets. One sheet was an article taken from Time magazine, July 12, 1999 and the other the assignment sheet. The original sheets are again presented in appendix VI. Similar to the *Dear Nobody test,* the Dutch assignment sheet has again been translated into English and is presented below.

Paragraph 1

1. In the text you will find a serious warning given by Prime Minister Tony Blair.
   a) Who is this warning for? (5 pts)
   b) What does he warn against?
2. 'No plan B to fall back on'- What does that mean? (10 pts)

Paragraph 2

3. What important condition, referred to as the most crucial basis for success of the peace process, is being discussed here? (10 pts)
4. What does reporter Helen Gibson think about the talks in this paragraph? (10 pts)

Paragraph 3

5. David Trimble and Gerry Adams are sent away to negotiate with their followers. Which of the two will, according to Helen Gibson, meet with the more resistance? (10 pts)

Paragraph 5

6. 'Decommissioning' – What does that mean? (10 pts)
7. David Trimble says: 'No guns, no government'. Gerry Adams says: 'No government, no guns'. What is the difference? (10 pts)

Paragraph 7

8. What is the 'seismic shift', as Blair so nicely puts it? (10 pts)
9. Why do Trimble's unionists object to the word 'could'? (5 pts)

Paragraph 10

10. 'The peace process suddenly looked…..' What does suddenly refer to? (5 pts)
11. What was the goal of the 'military operation' in Drumcree? (5 pts)

Paragraph 12

12. The article concludes that Northern-Ireland's future is in Trimble's hands. Explain why. (10 pts)

The reading comprehension test will again be discussed in terms of justification, knowledge, skills and insight measured, details on construction and use, and alleged washback.

8.7.1 Justification

Joy offered three reasons why she had selected this reading comprehension test.

- Her learners had on the whole done reasonably well on the test.
- There was a direct relationship between test preparation in which knowledge of and idioms related to the Irish question were acquired and the contents of the test. Joy considered this background knowledge essential.
- The prediction she had made before the test was administered about who would do well and who would do badly actually came true.

8.7.2 The knowledge, skills, and insights measured by the test

Again, Joy initially has difficulties in telling me what the test measures.

J: Erm, yes, pff. Well it's just like you said before. It really measures a lot of things. Initially it was meant as a reading comprehension test, you know, with open-ended questions. Try to put into words what this text is really all about, well, maybe not that so much as, erm, yes.. try to answer certain questions.

I: Could you specify the questions that measure reading comprehension? Types of open-ended questions?

J: Connections. But also just the actual meaning of a word in a certain context.

I: Sometimes also literal questions, that//

J: Oh yes. Although maybe those aren't really open-ended questions (laughter).

I: No, Okay, I agree.

J: But also about advance knowledge, or at least what we already mentioned in the course of the project, I mean.

I: So questions that are about logical connections between sentences or paragraphs?

J: Yes, between both.

I: Both. That are actually there. Sometimes literal questions about a certain word? [Yes] Perhaps antecedents as well.

J: Yes.

I: Any other things that you would like to mention concerning what the test actually measures?

J: Well.

I: Well, you don't have to. I meant what you would define as successful//

J: Well, maybe also the ability to summarize the information in a certain paragraph or something of the kind.

I: Are you thinking of certain key words that they should use, [Yes, oh yes, yes.] which you want to see back in the answer. In translated form, most of the time, in this case.

J: Yes, oh yes. Of course we opted to pose the questions in Dutch and to have them answer the questions in Dutch as well, because that is also the way it is done in the final exam. We discussed the issue, because during the practice phase it turned out, and this is usually also the case with the final exams, that it is very difficult for the learners to put the answer into the right words, to paraphrase adequately. And they do broadly know what the text is about, but, how do I have to express this, yes, what key words should I use, What//how do I extract these?

I: As concisely as possible.

J: Yes, as concisely as possible.
(4-1-70/71)

In the opening segment on knowledge, skills and insight, Joy first mentions the underlying skills of the ability to paraphrase what the text is about, which she weakens by adding ' try to answer certain questions'. Some of these questions assume the ability to see logical connections between sentences and paragraphs. Other questions depend on the learners' advance knowledge about the subject matter of the reading comprehension text. Some questions may also require the ability to summarise and/or paraphrase part of the text.

Follow-up questions invite Joy to define reading comprehension more specifically in terms of knowledge and skills (segments 72-74). The results of these specifications are presented below.

Knowledge

Joy mentions two knowledge areas. Learners need to have knowledge of:

1. the specific vocabulary and idioms required to understand and discuss the Irish question, and
2. primarily historical foreknowledge of the Irish question.

Joy refers to the vocabulary and background knowledge needed to understand and take part in English discourse related to the Irish Question. Joy tries hard to think of more knowledge elements, but cannot find any (4-1-73).

Skills

Learners 'should be able to read and understand a text and be able to concisely summarise what a particular passage is about' (4-1-74). Joy feels an important underlying ability is the appropriate recognition and interpretation of key words or notions that help a learner to understand what a text is about. Another underlying ability is the ability to see how sentences and paragraphs are logically connected in a text.

Insight

Joy feels that learners become more insightful readers if they master a certain amount of vocabulary, subject knowledge, and insight into how a particular language works (4-1-76). Joy arrived at defining insight in this way by reflecting on the achievements of one of her most insightful learners.

8.7.3 Construction and use of the *Irish Question* test

Construction (segments 4-1-81/82)

The first version of the test was constructed by a colleague of Joy's and Pete's who taught one 4-havo class. This colleague was used to teaching English to vmbo-classes. The first version of the test was commented on by Joy and Pete. Particularly Pete was not too pleased with the questions of the test and suggested questions that were more challenging and demanding. Some of his suggestions had been included in the final version of the test. It resulted in a compromise that Joy still finds wanting in a number of respects. We will mention her criticisms under the heading of assessment.

Expertise (4-1-86)

Joy acknowledges it is difficult to construct appropriate items for reading comprehension tests. The format and types of test items that are used in the final examination tests provided a standard for Joy to adopt when she constructed reading comprehension tests of her own.

> *J: A moral question, really. Effective items in reading comprehension tests. That is very difficult.*
>
> *I: It surely is.*
>
> *J: It is a matter of, well, what you sometimes ask of your learners. If you read a text and you wish to understand what it says, first try to formulate effective questions on that text. That's a logical starting point. And having a close look at the kinds of questions that are asked in the national exams, at what is expected of the learners.*
>
> *O: Yes, so CITO examples of test items of the final examinations help to//*
>
> *O: They do, for sure. Because we actually studied them.*
>
> *O: So, in the final examinations you come across items you find//*
>
> *D: Yes, also items that require learners to represent meaning, formulate or paraphrase in an economical and concise way. That is also why we scored the full points or no points at all, a procedure similar to the final examinations* (4-1-86).

Joy attempts to improve her own assessment practice by trying to formulate effective questions on the text and by analysing and reflecting on other tests. In that sense, the items of the national reading examinations had a positive washback effect on her ability to formulate test items with reading comprehension tests or on her ability to assess the quality of other reading comprehension tests.

Conditions (4-1-94)

The conditions of the test were unorthodox. For a period of fifty minutes, the text was given to the learners to study without the comprehension questions. The actual test was done in the next 50-minute class a couple of days later.

Joy comments on the approach as follows:

> *J: Yes, there was a break between the two lessons. A break of a few days even. They could have jotted down words and looked them up at home. No problem. They were even allowed to use dictionaries. And if they wanted to and they weren't sure what the Good Friday Agreement was all about, then they could have looked that up as well. So we just gave them the text, this is*

> *going to be the text, you are going to get questions about this text. And if you're smart, you might think of some questions yourself first. But anyway. They could also have discussed the text with others.*
> (3-3-94)

The approach is interesting. It gives learners the opportunity to get to understand the text and look up background information before the actual test is administered. Despite this procedure, four of her learners did badly on the test. with scores of 2, 2,8, 4 and 5,3. By contrast, six learners had scores between 8 and 9. The average of Joy's atheneum class was 6,3.

Preparation (4-1-91/93)

The test was again a summative test at the end of a project. The didactic components of this project were gathering and interpreting information on the Irish question, getting used to making and answering comprehension questions in Dutch on English reading texts, and a little training on how to make summaries of texts.

First, Joy provided information on the Irish question. She concentrated on historical facts and told her learners about the background of the present conflict. Next, the learners watched and commented on the documentary *Bloody Sunday.* This was followed by assignments related to the filmed version of *Cal,* a novel written by Bernard MacLaverty. Cal is a Belfast teenager who, against his will, is involved in the terrible war between Catholics and Protestants and is forced to make some devastatingly simple choices in his life: he can work in a nauseating slaughterhouse or join the dole queue; he can brood on his past or plan a future with the beautiful, widowed Marcella, for whose grief he shares more than a little responsibility.

An individual portfolio assignment was also part of the project. The learners had to collect materials on the Irish question in groups and build up their vocab files as preparation of a classroom role-play debate between Catholics and Protestants.

The project also comprised practice in making summaries and formulate and answer Dutch comprehension questions in terse and concise ways. Texts and questions were sometimes taken from the course materials *Touchdown*.

Assessment (4-1-79/81)

When we discussed assessment details, Joy offered her personal evaluation of the test. She was critical of the test, despite the fact that some of her learners succeeded in getting high scores. Originally, the idea was to assign 10 points for items that were supposed to be more difficult and 5 points to the items that were thought of as relatively easy. Joy felt that the division between 5-point and 10-point items was rather haphazard.

Joy also objects to items that are vaguely formulated, such as 10 'The Peace process suddenly looked as fragile-as as urgent- as ever'. The vagueness was in the question what 'suddenly' referred to. Joy feels the item can be interpreted in many ways. The response it was meant to elicit was that Gerry Adams was angered by Trimble's statement that Adams could not possibly guarantee disarmament. Item 12 was also difficult to score, because Joy felt she had to accept several answers. The result of the vague items was that 0,5 points would be added as a bonus to the final score for her 4 atheneum learners.

Joy also criticises that items 1-4 are easy and straightforward, and that from 5/6 onwards the items become more difficult and complicated. The test item she liked best was item 7. This was one of Pete's alterations, which he had suggested when commenting on the original test made by his colleague.

Evaluation (4-1-100/108)

Joy discussed the test results with the two colleagues who had also administered the test. It had not been a detailed evaluation of the test results. They primarily

discussed the results, and not so much the quality of the test. Joy spoke with Pete about the predictive quality of item 7. Learners really have to think and dive into the text to come up with an adequate response. Pete had constructed the item, without himself realizing that it discriminated so well between the learners who had prepared well for the test and those who had not. Both considered this an interesting test quality. Later, we will report that Pete's criticism of the *Irish question* test is downright devastating. We have already gone into Joy's main objections in the sections above.

Joy elaborately went into how the test was returned and discussed with her 22 atheneum learners. They always insist on a discussion of the test. However, the learners seem to be more interested in the actual results of the test, than in the learning process. Items 1-4 were not discussed. Hardly anyone had any mistakes in the first four relatively easy items. Items 5-12 were discussed in the way characteristic of Joy.

> *I: So in your classroom evaluation, you gave the correct response to a given item, which gives learners the opportunity to come up with suggestions such as 'Isn't my response just as correct'?*
>
> *J: Yes, exactly. But maybe, yes, it was a little wicked of me that I had said beforehand that the people who had apparently prepared well for the classroom debate, had no serious problems in answering most of the questions. This was probably the reason why the learners were hardly critical of the test. I would have expected the learners to object to the ambiguity and lack of clarity of items 10 and 12, but that didn't happen at all.*
>
> *I: No. The focus was really on whether their response was correct or incorrect. It makes me curious about how the learners assess the test.*
>
> *J: Me, too. Perhaps some learners will say that the test was simply too difficult in certain ways. I don't know*
>
> (4-1-101).

Joy tried to account for the learners' mood of resignation when the correct responses to the test items were discussed. She feels the learners have probably accepted the fact that good results were indeed possible if a lot energy had been put in the project that preceded the test, such as the preparations for the debate. However, the focus in this classroom evaluation is on whether a particular response is right or wrong, and why this is the case. It is product-oriented, instead of process-oriented. Joy is very much aware of this problem. The following extract links classroom evaluation to absence of test washback.

> *J: If you ask them, I think my learners are going to say they do very little with the tests that have been returned. Nevertheless, I keep hammering away at it. I just don't understand. I downright say, well, you have made some errors, which doesn't matter at all, but it is essential that you learn from the errors you have made, so that you do not make similar errors some other time.*
>
> *I: Yes, exactly.*
>
> *J: But for the learners a test is a kind of closed event. They have been preparing for the test for a while, and now they have finished it. And if they have scored a sufficient mark, well that's fine to them. What luck, now I will never ever need to look at the test again*
>
> (4-1-102)

Joy voices what she is up against as a teacher. She offers sound advice, even compels her learners to keep all the tests they have had in a file. Yet, it all seems to

no avail. Because this view of Joy's links evaluation to learner autonomy, I check whether I have understood what she meant to say.

> *I: Every test is seen as single event, which is unrelated to how good at English the learners feel they are.*
>
> *J: They don't see this as a process of some kind. They are just bits and pieces. Forever completed. For them it's all about the marks and if they are sufficient , you don't ever have to look at the tests again. The learners who are weak at English and would like to improve, they come and see you and ask for additional practice materials. But even they often forget to reflect on what exactly went wrong. That's really curious*
>
> (4-1-103)

The link between the evaluation of an assessment procedure between the learner and the teacher as assessor and the learner's willingness and ability to learn, will be discussed in our final chapter. At this point it seems fair to remark that beneficial washback effects on the ways in which the teachers teach and the learners learn seem hard to achieve if evaluation is primarily product-oriented.

We will finally report on any washback effects on communicative language education or learner autonomy of the *Dear Nobody* reading comprehension test.

8.7.4 Washback on CLE

Joy again feels that the test itself does not stimulate positive washback on learning how to learn to communicate in English. But, as was the case with the *Dear Nobody* writing test, she claims that the project on the whole had stimulated how people learn how to communicate in English. The fact that the summative test focussed on reading comprehension was partly because the programme of assessment and summative evaluation (PTA) stressed the need for assessment of the learners' reading comprehension. It could have been a different test, but reading comprehension needed to be tested at this particular stage. (3-3-109). Joy seems to sense that communicative competence is made up of a mix of knowledge and skills that are related to one another in complex ways. Seen in that light, she thinks it is odd that the four language skills need to be assessed separately and are rarely ever seen, let alone, assessed in combination.

We already know that Joy objects to a discrete-point approach of foreign language education. She feels it is important to acknowledge that in authentic foreign language use, knowledge and skills are always present in integrated ways. That is why she feels that the traditional approach of testing reading, writing, listening and speaking is at times troublesome and confuses her foreign language learners.

> *J: But what we do find very annoying is that you were pretty much forced to develop this file incorporating the four language skills and that was difficult. We have these integrated projects, so when you watch the movie, "Cal', and you take notes and you talk about the movie and then your learners ask well, what goes under 'watching' and what goes under 'speaking' and what goes under 'literature' or something. Yes, then I thought, but of course that will just drive you nuts.*
>
> (3-3-108)

Joy feels that in an effective project the skills should always be integrated. This had also been the case with the *Cal* project. Besides reading comprehension, other knowledge and skills that were part of the project were in the PTA for 4-atheneum. The classroom role-play debate between Catholics and Protestants was seen as a so-called "handelingsdeel" for fluency, the novel *Cal* will be dealt with in literature classes, note-taking was an important strategy that was practised to have the

learners develop their listening skills. Besides, the project had extensive input of authentic reading texts, partly provided by the learners themselves in their portfolio on the Irish question. Therefore, tests on such projects that are to have both content and construct validity need to assess more than one particular language skill that has been singled out for the sake of convenience.

Let us finally consider how Joy related the test to learner independence.

8.7.5 Washback on LA

Joy had already informed us about the positive effects on learners who had actively participated in the project and prepared the individual tasks well. They also appeared to be the learners with high scores on the reading comprehension test. Moreover, the reading comprehension text was given to the learners to study at home before the actual test was administered. It is therefore likely that this particular way of assessment had a positive effect on the learners who scored well. Yet, conclusions such as these are all at an interpretative level. Washback does not seem to be on the minds of the learners. Neither does Joy seem to go any further than strongly advising her learner to reflect on the tests and relate them to their English proficiency.

We were more than a little interested to have Joy probe more deeply into the heart of problems with test washback as she saw them. We asked her what actions she would attempt to undertake to stimulate learners to look back on mistakes and errors made in past tests to prepare for tests to come.

> *I: And what do you do about that as a teacher? How do you try to//*
>
> *J: Yes, try. I try to focus on it all the time, but on the other hand, when I think back to when I was their age, then I didn't do that either.*
>
> *I: But you did have passing marks for English, didn't you?*
>
> *J: Yes. Although I do remember one time, in the fourth form, I had always passed my English tests, but then we got this new teacher, and the focus was also on grammar and the tenses, and then I did suddenly get a three back. And then we, all the learners had bad marks, so then she said, well, I am not registering this mark, because with so many bad marks let's first get some work done. But do you think that we all tried to find out what we had done wrong? We just thought, this is really a bad test. Way too difficult. But we never thought to look at our mistakes. I never did. It was just done, finito. I had other things on my mind. Was busy doing other things. It's actually very difficult to try to change that cycle. … But I think that learners, at that age, they're just really focused on the short term. You don't really consider the long term effects. I've noticed that it's really difficult to break through that. It has everything to do with their own goals. Yes, they're thinking about what do I want to be when I grow up, but they're mainly thinking about the short term. I have to get a good mark for that test. And that's final.*
>
> (3-3-104)

Joy refers to her own experiences as an adolescent learner, which leads to the view that most adolescents have other things on their minds than error correction or any positive effects derived from bad scores on a test. Joy acknowledges that learners see tests as summative assessments of past achievements that hardly bear any relation with knowledge and skills to be tested in the future. She tends to accept the phenomenon as a given fact. She states the issue, but does not see any solutions readily available.

The key element in Joy's feedback on the results of language tests is the sound advice she gives to her learners, to which extra materials are added if need be. Joy feels that there are limits to what a teacher can do and that learner autonomy

ultimately is the responsibility of the individual learner. What Joy at best hopes is that learners realise that effective preparation to a reading comprehension test of this kind pays off by way of good test results (4-1-110).

We ended our last interview with another issue. Joy's stated that she felt the mastery of vocabulary of her classes in the second phase was disappointingly low, and that they were thinking of more explicit vocabulary teaching in the school year to come. According to Joy, all of the extensive reading and listening tasks of the projects and the learners building their own personal word files had not resulted in the mastery of vocabulary needed to do well on the present and future tests and examinations.

In the next section, we will relate the data from the tests and interviews to Joy's core beliefs and construct definitions.

8.8 Joy's core beliefs and her assessment and evaluation practice

The first analyses of the interview data soon made clear how important the role is of a teacher's core beliefs in relation to his/her evaluation practice. Evidence of Joy's professional views in her language tests appeared to be converging in a convincing way. In this section we will report on this evidence by mentioning how her convictions are reflected in her testing practice. We will do so by first referring to three important general characteristics of Joy's assessment and evaluation practice. After that, we will more specifically discuss each of her core beliefs in relation to the analyses of the tests and the ensuing interviews on them.

8.8.1 General characteristics

All of the three tests selected by Joy provide convergent evidence of her core beliefs in the sense that:

- two of the tests are based on and are strongly related to thematic projects that include a variety of communicative individual and group tasks;
- Joy prefers integrative knowledge and skills to discrete-point knowledge or skills, and therefore likewise prefers integrative tests to discrete-point tests. The first and third test she had selected for discussion were summative tests at the end of a project. Despite the fact the two tests seem to focus on one language skill, i.e. writing or reading, the tests seem to require the integration of general and subject-specific knowledge and skills that were built up in the course of the various tasks and activities the learners were asked to carry out. The focus on a particular language skill seems mainly determined by the skill that was required to be tested in the PTA Joy helped to construct (Programme for Assessment and Summative Evaluation).
- the only test in which grammar was tested focused on verbal tense and form and was constructed by changing the verbal tenses and/or forms of an authentic text into infinitives. Test preparation, again, involved a teacher introduction, individual work and group work.

8.8.2 Core beliefs in the tests and interview data

We see that the three tests have provided convergent evidence of twelve of the fourteen core beliefs she expressed in the first interview.

1 *the advanced learner and/or beginning teacher developing insight in how the foreign language works the moment she actually starts teaching (1-1-27);*

Evidence of this particular belief of Joy's is less strong than is the case with some of the other core beliefs we will discuss below. In general, there is little direct evidence of Joy having the learners focus on the process of language learning, rather than on the actual products. The most concrete example is when she has her learners work in pairs and small groups to explain to one another what particular tense and/or form has to be used when the class was preparing for the Little Boy grammar test. Process orientation is certainly on Joy's mind, but does not seem to have developed any further than sound advice she offers the learners to look once more at their errors and/or mistakes of all of the tests that are kept in the test files. In this sense, Joy truly is a budding professional.

2 *the benefits of exploring literary texts: the main reason why she decided to study English (grammar and fluency she considered her weaker points);*

In the two projects on which two of the tests have been based, Berlie Doherty's Dear Nobody and Roddy Doyle's The Snapper played important roles. In both instances, extracts from the novels have been combined with sections from their filmed versions.
However, the literary exploration of the two texts appears to have been limited to providing readable, attractive and authentic alternatives to other texts on the themes of the projects. Implicitly, such a focus on literary texts may lead to increased appreciation of literary texts and a better understanding of the language that has been used.

3 *teachers developing their own materials and tests based on or taken from authentic texts about themes that appeal to the learners. Joy strongly believes in meaningful texts. (1-1-12);*

Evidence about this particular belief appears to be convincing: all of the authentic texts and materials that had been used were part of clearly defined and well-prepared thematic projects that appeared to appeal to the learners.

4 *teachers cooperating closely, e.g. on what to test at what particular time: exchanging and discussing ideas are motivating experiences for the teacher. (1-1-16);*

As she had indicated in the first interview, Joy cooperates closely with Pete. Both teachers motivate one another. Pete's creativity, speed and slight disorganization, combines well with Joy's creativity, deliberation and her talent for organising education in adequate ways.
However, the only colleague who also teaches fourth forms does not seem to share Joy's and Pete's ideas about and skills with effective teaching and testing. It forces them to occasionally compromise, as is shown by the Irish Question reading comprehension test.

5 *tests that call for more knowledge and skills than mere short-term reproduction of discrete elements that have been studied by heart. (1-1-2);*

Again, evidence appears to be overwhelming. Neither the tests, nor the tasks that led up to the tests focused on mere short-term reproduction of knowledge. Joy's intention is to always have her learners use and/or interpret the knowledge and skills that have been acquired in the context of meaningful English texts when they are assessed.
However, Joy appears to have difficulties in identifying the knowledge and skills the learners need to successfully do an integrative test. The result is that the focus is more on the actual product than on the conscious building of the microskills required to carry out communicative tasks.

6 *the uselessness of pattern drills and the reproduction of standard sentences in teaching learners how to learn to communicate in English (1-1-31);*

Pattern drills do not at all seem to play a role in Joy's teaching. Standard sentences, however, do to a modest extent. When the learners are first introduced to the heuristic 'when-scheme', this is first of all practised by way of standard translation sentences, in which time adverbials are easily identifiable and the verbal tenses and/or forms that have to be used fit in with the rules of the scheme. Yet, Joy soon replaces these standard sentences by working with authentic texts of increasing complexity in terms of verbal structure and vocabulary.

7 *the uselessness of teaching the transformations needed to turn active sentences into passive ones; recognition of passive constructions is enough for her. Energy is better spent on teaching learners how to use proper tenses and verbal forms. (1-1-11);*

The focus on verbal tense and form has been illustrated by the gapfill Little Boy test at the beginning of the school year. This test and the exercises leading up to it is referred to if verbal errors and/or mistakes occur in the course of the school year.

8 *the uselessness of rote learning and reproductive testing of idioms and her choice to test idioms in context of authentic written tests or by way of writing assignments. (1-1-18);*

Idioms have not been tested explicitly in the tests Joy offered for discussion. In all the tests, the meaning of the texts was either assumed or the learners were given the opportunity to look up idioms before the test. In the final interview, Joy claims to regret such a focus on implicit idiom teaching, because she feels her second phase learners master less vocabulary than their predecessors in previous years. We will return to the role of vocabulary acquisition in our discussion chapter.

9 *long-term retention of vocabulary, e.g. by having the learners set up their own word files and immersing readers in extensive reading of authentic texts that are slightly more difficult than what they are used to handling.(1-1-6);*

As indicated above, having the learners set up their own word files and confronting them with extensive reading and/or listening texts has not resulted in knowledge of vocabulary that Joy considers as adequate for fourth formers. If Joy is right, this may have serious consequences in the years to come and may lead to a return to having the learners explicitly study idiom lists and do a variety of explicit vocabulary exercises. A drawback is that Joy herself speaks very little English in class and that English is not required when reading comprehension is tested.

10 *integrative tests rather than discrete-point tests (1-1-6);*

In all of the tests, several types of knowledge and/or skills have to be integrated in order to successfully do a test. As stated before, the more complex this knowledge or skill is, the more difficult it is for Joy to indicate what is being assessed in a precise way. It actually was one of the reasons for discussing the Little Boy grammar test, with its knowledge and skills that were more easily identifiable. Yet, also this test was intended to require more than the application of the 'when scheme' by offering the verbal structures in a structured and authentic meaningful text.

11 *the importance of reassuring learners who are likely to panic. (1-1-7);*

Joy always discusses the intended response of a test after it has been returned and always gives her learners the opportunity to ask questions on it. Joy does not mention any learners having panicked about any of the three tests that have been discussed. However, some learners who scored badly were eager to do some extra work, which Joy readily supplied.

12 *the fact that reliably and validly assessing writing tests is problematic (1-1-23);*

Writing is a productive skill that does indeed require a variety of knowledge and microskills. Joy gets no further than roughly distinguishing between the 'technical' part and the 'content' part. Reliability, however, is on her mind, because she discusses any response she is not sure of with Pete.

13 *the motivating effects of offering the learners some choice, also in areas they are not intrinsically interested in, e.g. literary texts (1-1-34)*

The element of choice has only been explicit in the Dear nobody test, where the learners were free to choose from one of the three writing assignments. Joy attaches great importance to selecting project themes and activities that are likely to interest the young learners.

Above, we have found convergent evidence of thirteen of Joy's beliefs. No discriminant evidence of her core beliefs was found in the course of the school year, except for the remarkable absence of Joy speaking English as a classroom language. Yet, there is one of Joy's beliefs we found little evidence of. We will briefly discuss this before turning a concluding summary of the findings of the three follow-up interviews.

At the beginning of the school year, Joy stated that she wished to make "how a language works more of a focal point in her teaching and testing practice" (1-1-27). Despite her efforts to focus on language learning processes next to the results of these, there appears to be limited reflection on what exactly is needed to become a proficient user of English in terms of knowledge and skills. It seems hard for Joy to be explicit about these, especially when it concerns a productive skill such as writing. Yet, her mind is set on changing this, even if she has not succeeded in developing this belief of hers in a more specific direction.

8.9 Summary

In this chapter we started with an overview Joy's core beliefs and construct definitions as expressed in the first interview, which had been presented in chapter 7. We then focused on the three tests that Joy had selected for discussion in the course of the school year. The first test we went into was the *Dear Nobody* writing test. The writing test was a summative achievement test on what had been learned in the course of a two-month project on teenage pregnancies. The project itself was based on Berlie Doherty's novel *Dear Nobody* and also used extracts from the filmed version of Roddy Doyle's *The Snapper*. The test was illustrative of Joy's profound beliefs in projects, in which the language skills have to be used integratively in meaningful ways. The second test we discussed was a grammar test that had been administered some six weeks after the start of the school year. It was based on an authentic text taken from the *Daily Mail* , which was entitled *Little Boy Growing old before his Time*. The text had been changed into a gapfill text by replacing thirty verbal tenses and/or forms by their infinitives. The learners had to replace the original tenses and forms with the help of implicit and explicit information in the text. The *Little boy* grammar test illustrated Joy's view of the role of grammar in learning how

to learn to communicate in English. It also illustrates her focus on a limited number of essential grammar points, such as verbal tense and form, or word order. The final test Joy reflected on was the *Irish question* reading comprehension test. It again was a summative achievement test that concluded a project. The test had been constructed by a colleague of Joy's and Pete's, with some changes added by Pete.

All of the three tests that Joy had selected were discussed in terms of Joy's *justification* why the particular test had been selected for a more detailed discussion. Then the tests were discussed in terms of the *knowledge, skills and/or insights* the tests were supposed to assess. This was followed by a more comprehensive section on the construction and use of each of the three tests. Information was presented on how the test was constructed, the knowledge and skills the teacher required to construct the test, the ways in which the learners had been prepared to the test, the test conditions, details on the ways in which the test was scored and graded, and finally the teacher's expectations of what learners (should) do with the test after it has been returned. The final two sections respectively concentrated on w*ashback* on *Communicative language Education (CLE)* and on *washback on Learner autonomy (LA),* in which Joy explained in how far the preparation, administration and discussion of the test stimulate the learners to learn more independently and responsibly.

In chapter 11 we will summarise the data from this chapter and compare and contrast this with the data from the chapters on Mark and Pete. Our findings will then be interpreted in view of the contents of the three theoretical chapters of this study in a process generally referred to as analytical generalisation.

Our next chapter focuses on Mark's language tests.

CHAPTER 9: MARK, THE LITERARY MASTER

9.1 Introduction

This chapter on Mark's tests has a structure similar to the chapter on Joy's tests. We will again start with a summary of the respondent's core beliefs and construct definitions as they were expressed in the first interview. Then we will briefly introduce the three language tests that Mark selected for discussion. In the three sections that follow, we will discuss the language tests under the headings *justification*, *conditions, knowledge, skills, and understandings* measured by the test, ways in which the learners have been *prepared* to the test, details on the *construction and use* of the test, and finally, details on the ways in which the test was *assessed* and *evaluated*. We will end our chapter on Mark with a discussion of Mark's core beliefs in relation to his assessment and evaluation practice and a summary, in which we will recapitulate the contents of this chapter.

9.2 Mark's beliefs and construct interpretations expressed in the first interview

As illustrated in chapter 7, Mark believes in:

- the value of teaching about literature and analysing literary texts. To this end, Mark wishes to empower his learners to be able to analyse and discuss poems, short stories and novels. He is convinced that by reading literature, one learns about life;
- the following main educational goal: 'to have learners leave this school with a fair amount of schooling in literature, next to having acquired quite a fair amount of knowledge and quite a fair knowledge of the English language as a means of communication, both oral and written.'
- the benefits of ex-cathedra, direct teaching, where an educator inspires his learners in convincing performances;
- the teacher setting examples of behaviour, linguistic or otherwise, for the learners to copy;
- transferring the academic approach of dealing with literature he had experienced himself to the English lessons at his grammar school;
- a division between junior secondary education (forms 1-3), in which a firm foundation is laid of grammar and vocabulary, and upper secondary education (forms 4-6), in which that groundwork is extended and transferred to a variety of literary texts and tasks;
- explicit grammar and idiom teaching at a highest possible level, regularly tested in translation sentences from Dutch into English;
- gradually and systematically building up a learner's literary and linguistic knowledge and skills over the years;
- the importance of contact hours in teaching and coaching his learners;
- the freedom education used to offer him as a teacher, which sharply contrasts with the rigidity of second phase innovations and the egalitarian nature of basic secondary education;
- the failure of the recent curricular and didactic innovations of basic secondary education and the second phase. He feels both innovations are responsible for two transitions he very much regrets: a move from teaching idiom and grammar from junior secondary education to upper secondary education and the diminishing role of the teacher, because learners are expected to work on their own most of the time;

- the ambivalent nature of course materials: on the one hand they are essential in structuring and shaping education, on the other the contents and assignments generally do not do justice to the grammar school learner;
- the shortcomings of tests that come with the course materials, which often have to be adapted by the teacher;
- the benefits of error correction and having the learners analyse the errors that have been made;
- the assessment of the learners' proficiency at the end of the learning process, and not during the process of learning to communicate itself;
- the incentives that tests provide for human beings in order to get work done and achieve results;
- the uselessness of the formal tests that he was forced to administer to his junior secondary learners when they closed off basic secondary education. Mark claims he has never administered these;
- the fact that the scores on the national CITO reading comprehension tests are overrated in the final mark a learner gets on his school certificate;
- the fact that the role of literature has been marginalised in the secondary phase innovations.

Mark's construct definitions

1. An effective written English language test:

- relates to the level the learners have attained;
- contains test tasks that help to assess in how far the learners are able to creatively use the grammar points and idioms that have been studied, practised and learned.

2. Essential knowledge and/or skills in English

- Knowledge and skills are largely determined by the contents and gradation of the course materials Mark has opted for and uses;
- *Spelling* and *formulation*, which are indicators of a learner's mastery of appropriate style;
- *Literary insight*: ability to justify personal opinions on the contents of literary texts;
- *Grammatical insight*: modest role of learning about grammatical and/or semantic relationships between the words and phrases in an English sentence.

3. CLE

- Communication is seen as an essential life skill. Teachers teach learners how to learn to communicate by setting examples, direct instruction and organising practice, the level of which has been determined by what the teacher is expected to teach at a given stage in the learning process. After that the learners are equipped to practise with and learn from their peers;
- Mark feels grammar teaching and developing a learner's language awareness are important components when teaching learners to communicate in English, e.g. by asking learners to explain the differences between sentences such as 'This is John's picture; This is a picture from John; This is a picture of John's ; This is a picture of John.'

4. LA

- Mark feels the construct *self-regulated learning* involves three competences:
  - *planning skills* and the discipline to realise what has been planned;
  - *self-assessment skills* allowing he learners to evaluate their own knowledge and skills in English;
  - the ability to *balance learning and doing*, e.g. the ability to distinguish between explicit learning activities, such as rote learning or applying learning strategies, and doing exercises.

- Mark is of opinion that the teacher's role is very crucial in having the learners assume responsibility for their own learning. It often involves intensive mentoring and coaching. At certain stages it may be necessary for the teacher to act and (re-)establish the conditions needed for learners to develop their autonomy.

5. Washback of tests on learners learning how to communicate in English

- Tests provide incentives for human beings to learn and achieve results. The results of tests enable both the teacher and the learners to reflect on the results that were expected before the test was done and the actual results that were achieved. Both the learner and the teacher can use the results to evaluate teaching and learning.
- Nothing succeeds like success: good marks enhance learner interest and motivation.

9.3 Mark's tests and interviews

The three tests that Mark had selected for discussion considerably differed from the tests Joy had opted for. As was the case with Joy, the first test that Mark wished to discuss was most indicative of his core beliefs. It was a literature test on Hemingway's short story *A Day's Wait.* The second test is an instance of a type of test his learners had to take every five weeks in a school week reserved for assessment and other activities. It was a regular course materials test that tested essential knowledge and skills of a unit of *Unicom Finals* that had been dealt with in class. The final test Mark had opted for was again a literature test**.** In this test a group of three or four learners had to discuss and write an essay about a novel of their preference. This test was originally part of Mark's PTA as a so-called *handelingsdeel*, or dossier test. In a dossier test, a learner has to provide evidence of mastering certain skills or competencies in a comprehensive test task. Such tests are usually only scored with a pass or fail. However, in this case scores between were given for the group essay on a ten-point scale. This meant that Mark had to rename the test in his PTA. The test was officially called a *practical assignment*, so that marks could be given.

All of tests were informal tests. The literature tests had been constructed by Mark himself. The course materials tests had been taken from the course materials, allegedly with slight adaptations and the addition of a text with reading comprehension questions.

We will now discuss Mark's first test in more detail.

9.4 Test 1: The *A Day's Wait* literature test

The first test illustrates Mark's core beliefs and concerns in English language teaching and testing. It was a summative achievement test his learners took after having been introduced to a number of literary terms and the type of discourse Mark considered essential for his learners to learn. Thus, the learners had become familiar with terms such as *theme, character, story line,* or *construction of the story.* The learners had practised answering questions on a number of short stories in class. The knowledge and skills were now tested by way of a short story they had not dealt with before. The text and test can be found in appendix VI.

9.4.1 Justification

Mark had chosen for the *A Day's Wait* test for three reasons.

- The test was chosen because Mark had used it before with success, as experienced by himself and, allegedly, most of his learners (2-2-1);
- *A Day's Wait* illustrates an important stage of Marks's didactic approach to literature, where he has his fourth formers master a number of literary notions and apply these notions to a written analysis and discussion of a short story (2-2-1);
- This particular test had been chosen, because it fitted in well with the test criteria he had in mind:
  - A concise story that appealed to the readers;
  - A test to be administered in a 50-minute lesson. (2-2-1)

9.4.2 The knowledge, skills, and insights measured by the test

This is how Mark responded to our question what exactly was measured by the test.

> *I think they are intertwined, but of course it measures if people have understood the theme and what is meant by 'a character'. What is the 'construction of a story' and such things? How do you make a 'summary', and 'outline'? Because that isn't very easy. That in the first place. Then in the second place, I think, and that often becomes clear when you look at their answers, you're also testing the stage of the learners' development, but only to a limited extent I would like to stress. You can say that literature is often about life. And if literature appeals to the learners' imaginative powers, because they themselves share experiences, or witnessed them, or have already read about them, you will notice that their development in grasping stories, in understanding relationships, is at a higher level. On the other hand, though, you also have children who are really, well, maybe because of their personalities, or the way in which they were raised, but also because they are just still very naïve about life, or just think differently, and they give answers that show that they understand what it means, but they just don't go for that deeper understanding. They don't have the feel for it, YET (emphasised).*
> (2-2-5)

According to Mark the short story test assesses:

- whether the learners have understood the literary notions dealt with in class, such as 'theme', 'characterization', and 'construction of the story';
- whether the learners can produce adequate outlines or summaries of a short story;
- the stage of personal and moral development of his learners to a limited extent.

Mark feels literature is about life. Learners differ in their imaginative powers or capacities to identify with matters that are explored in literary texts. There are learners who have not yet been able to arrive at a deeper understanding of literature. Such learners have difficulties in understanding what *A Day's Wait* is about. Identification with its theme, plot or characters will be difficult. By way of example, Mark refers to a question asked by a learner who had scored badly on *A Day's Wait* and failed to understand why. The learner felt he had answered the questions on the boy's 'character' correctly.

> *Well, yes. I have this extreme example of this boy who afterwards, after receiving a disappointing mark, for both him and myself really, asked me "Yes, but sir, didn't I get everything right?" But I remember that he answered an item about the boy's personality by providing specific details about his age, his size, and he went to France and went to school there. But he never ventured inside the character, you know?*
> (2-2-5)

Mark subsequently points at a dilemma of literature education, which at the same time stresses the need for learners to have their horizons widened by way of literary texts.

> *(If you) present more literary texts to children who are 17 /18, how far have they come? You' re really giving them the key to a literary room, or literary house, in the hope that later, when they're ready for it, they will get something more out of it, you know. I had someone in the fifth form who had read Sophie's Choice for the 'war project'. Well, of course she reads the novel in a certain way. And she extracts a lot of meaning from it, but she will also have missed a lot of information. And I don't mean to speak badly of the girl. But really, it's all about still being young, isn't it? Just not understanding everything yet.*
> (2-2-5)

Mark stresses that the gains of studying literature are immaterial. These benefits are at odds with today's tendency to discuss literature education in terms of its use. Marks feels what should be discussed and appreciated is its indispensable value in learning about life and stimulating independent thought in that process (2-2-6).

Assessing the degree of identification and personal or moral development of a learner is no easy matter. That is why generally Mark prefers oral testing of literature to written testing. Oral testing involves teacher-learner interaction, enabling the teacher to put a learner on the right track if (s)he misinterprets a question. It is important for Mark to notice in how far his learners have become personally engaged in a short story or novel, and have attempted to mirror the experiences and feelings presented there with their own (2-2-7). He feels it is important to stimulate the learners by giving positive assessments.

> *Well, yes, I enjoy giving positive assessments. I also find that in literature, you should appreciate what comes out of them, not what they don't understand. Other than the odd cases who have only read summaries, and didn't read the book or didn't put up the effort. Because it should really be a stimulating process. That's a fairly romantic view to hold in this day and age, but that's just my// and I'm the kind of person who says, show me what you can do, you know. I find a question about the romantic elements of a literary text quite interesting, and if you can do it, that's fine, but it's not the point of departure. You start with personal experiences.*
> (2-2-7)

Despite the fact that Mark claims that the learners' personal experiences are his point of departure and that widening and deepening these experiences is his ultimate aim, he attaches great value to knowledge about literature and literary terms. When we discuss the knowledge, skills, and insight measured in the *A Day's Wait* test, Mark's answers come straightforward and with relative ease. Despite this ease, he says he finds it hard to distinguish between knowledge, skills and insights. First, he feels one cannot separate knowledge elements entirely from skills.

This is how Mark responded to our question on what knowledge, skills and/or insights were measured by the test.

M: Yes, what they have to know and be able to do in order to do well on the test. Well, question 1 is really a practical assignment, that// you can work on it and you can improve it, you know. A brief outline of only the backbone of the story. And that's what we practised over and over again, because writing the shortest possible summary of the story proved to be the assignment that was most difficult. I often indicated, well, five sentences. Or six sentences, and then I would call on some people in class and then I would ask the rest, hey, what could you leave out here, or what's missing here? So a – well a, a specific skill, combined with knowledge. Because skills without knowledge, well that's just rubbish of course.

*When you're speaking of themes (**2**), then you're in a rather grey area, I think, because people might know what a theme is, theoretically, shall I say, but they could still be way off the mark. They still have the tendency to sometimes do that with this story. The question about the boy's character **(3)**, again that has to do with, well, how do you relate to him, how far have you got yet, you know? A little personal insight into this person, although they have also been taught what things to consider. And the same goes for the question about the relationship between father and son **(4)**. Question five, the crucial line of the story **(5)**, is basically a question that assesses their insight. To see if they have understood the story, yes. And the same goes for question six **(6),** which just goes to show that a lot of the questions are related. And if you were to do this test orally, you could steer people in the right direction when they are dangerously close to taking the wrong turn. Then it would be a better assessment tool, because otherwise you tend to, almost like a mathematician, you know, WRONG, WRONG (emphasized), because those steps. Yes, well. Time is often the – the major factor that …*

I: Because of which you simply can't do this test orally?

*M: No, no, you could hardly do it. Well, the same goes for question seven **(7)**. That's also more of a comprehension question, isn't it. It also kind of relates to the ability to empathise. Question eight **(8)** is more objective, you might say. It measures, you know, the construction of the story line. You can really point that out and explain why.*

(2-2-8)

I: They should be able to discern the three parts? [yes] Yes.

*M: Question 8b was a lot more difficult, it appeared. But anyway, we'll get to that … [yes, definitely]. The title **(9)**, that once again boils down to interpretation, you know. It has to do with the theme and with … And question ten **(10)**, you could always score ten points for that. Even if all you wrote down was "I like the story", although I did ask them to provide a bit more elaborate an answer than just that. But some people ran out of time and that is basically the sentence you start with. Often I will start with, also with an oral exam, with what did you think of the book? Or what did you think of the main character, or Holden Caulfield, when you're discussing 'Catcher in the Rye'? "Could he be your friend? And why?" [Just to get their own opinion and then…] Yes,*

> *and ultimately you arrive at – and make sure that you have arguments that justify your opinion. [yes] That's really what I always try to ---. [yes]*
> (2-2-9)

According to Mark, knowledge always forms the basis of skills. Second, he feels that most of the test items measure, 'comprehension', 'insight' or 'interpretation'. Mark divides the ten test items of *A Day's Wait* into four categories:

a. **1** (Outline) & **8a** (Division into three parts): knowledge and/or skills that can be assessed technically and objectively;
b. **3-7** & **8b**: test items that Mark seems to interchangeably label as measuring 'comprehension', 'insight' or 'interpretation'. What these items seem to have in common is that they require some form of higher-order thinking skills;
c. **2** (Theme)& **9** (Explain the title) are labelled as 'hit-or-miss questions'(2-2-24);
d. **10** (Personal valuation): an important question for Mark for which he has awarded the maximum score of 10 points for each and every learner

When Mark came to discuss the ways in which he had scored and graded each and every item, he was admirably detailed and precise. Mark is a teacher who knows what he is after. He rarely has any doubts about what the required response to the test items should be. We will return to some of Mark's detailed answers when we discuss how Mark has scored and graded *A Day's Wait*.

9.4.3 Construction and use of the *A Day's Wait* test

We will again first deal with general information on the *construction* of the test. Then we will present the *expertise* Mark feels that is needed in order to construct the test. This will be followed by details on the ways in which the learners have (been) *prepared* to the test. Finally, there will be a brief section on the test *conditions* and longer sections on *assessment* and *evaluation.*

Construction (2-1-23)

Mark constructed the test himself and has used it for a long time. The copies Mark handed over, were copies of an original that had once been made with the help of a good old typewriter and duplicating machine. The original of the test is presented as an appendix. The copies handed out to the learners also appeared to have been around for a while.

A Day's Wait is part of a file of short stories and questions that Mark has constructed since he started his teaching career. The test is used as a preparation to more elaborate future discussions of novels. Besides, the test is used to give feedback on how the learners can interpret a short story using the terms and discourse that have been dealt with in class.

Expertise (2-2-23/24)

Mark says that the knowledge and skills needed to construct a test like this were gradually built up in his teaching practice. At first he used materials and tests that were already available at his grammar school. When he started teaching upper secondary classes, he initially dealt with stories he liked and gradually started building up the skills required to use literary texts, adapt them if need be, and construct useful test questions.

> *But in my first year of working here, yes, you could use this. If you explain this and perhaps also have a look at the correct answers, because, of course, you're rather*

insecure. The first stories I treated off-head. They were stories that I myself liked. So then you do Roald Dahl, and things like that. Yes, as time goes on, you're beginning to feel a little braver. Teaching literature is a subject in itself. It can be very difficult in a way. But captivating as well.
(2-2-24)

A Day's Wait has been used for such a long time that Mark cannot exactly remember how he has acquired the skills to construct a test like this. He is sure that the skill of selecting and adapting literary texts and constructing literary tests was not something Mark learnt at university.

M: Yes, I learned that a long time ago. And – yes – how do I do those things? I don't know. My grammar, for example, well that's all up here. (points at his head) That's been stored there. I never wrote it out. Sometimes I'll say, look at the notes of a good learner in the first, second, or third year. Well, then you've got a beautiful review. But yes, you have certain terms in your head, you know? Theme, for example. Well, that has to be covered and, yes, you give it a place in certain questions. [yes, yes]

I: Is this in any way related to your own academic studies? That you were used to handling certain literary notions as stepping-stones to discuss literary texts?

M: Well, that's a long time ago, isn't it, a long time ago. I think my stay at the, say, Alma Mater, did not lead to a lot of skills relevant to teaching literary texts. The focus primarily was on knowledge and the introduction to the vast field of literature. And I think that some lecturers were more successful than others, you know. But making literature, or stories to be more precise, practicable (with emphasis) I have only learned in everyday practice. And, adding to that, this is really is something that can be very useful for a host of beginning teachers as well. Not only the grammar, because that's what it often concerns, but also and very specifically a short story, with test questions, like the one we are talking about now. They don't have to copy this, but it gives them something to hold on to, you know, an example of how it can be done. And at the beginning they are likely to literally copy those stories as well.
(2-2-24)

In no uncertain terms, Mark expresses his belief in beginning teachers initially copying the teaching and assessment practices of more experienced teachers.

Conditions (2-2-24)

A Day's Wait was administered in a 50- minute lesson. Both the test items and the learners' response were in English. The use of dictionaries was not allowed.

Either a Dutch translation or an English synonym was given for five idioms: 'varnished with ice' (Du: met een vernislaagje van ijs); 'a covey of quail' (a flock of birds (kwartels), 'brush' (Du: kreupelhout), 'to flush' (to fly up suddenly), and 'poised' (balanced).

Every five lines of the short story had been numbered for ease of reference.

Preparation (2-2-1, 2, 3, 4 &10)

From the first form onwards, Mark's learners have been used to writing book reviews. Over the years the complexity of the reviews increases.

In the fourth form, Mark first deals with a number of literary notions such as 'Theme, setting, characterization, personal valuation, atmosphere, tone of voice. Et cetera. Style.' (2-2-1). These notions are then applied to four or five short stories he deals with in class. The tasks and activities involved with these short stories echo the

test items the learners will have to respond to in the summative test about a short story they have never seen before.

The short story for the test should literally be short, as it is meant to be administered in a 50-minute test.

A particular skill Mark first demonstrates and then has his learners practise is how to write effective outlines or summaries. The learners are taught to present outlines in five or six lines, based on key words. Coached by Mark and by comparison of a learner's summary with those of his/her peers, the learners are taught how to produce effective and adequate summaries.

All of the literature lessons in anticipation of *A Day's Wait* were meant to ensure that the test items would not come as a surprise to the learners, in other words that the test was to have face as well as content validity.

Teacher direction and a little coercion if required are characteristic of how Mark prepares his learners to the test.

> *M: Well, I begin by writing on the board the rather dry literary notions to make clear that literature is not just about enjoying a story, but that you must be able to explain why, and in addition appreciate it as a work of art. (In English:) 'If I can make you appreciate art, that's OK, but if I can make you like it, it's a bonus, isn't it' (Marks resumes to Dutch), I often say, because that of course is the idea behind it. To put them on the track of a number of interesting authors, and to have them read a story or novel more often. And yes, this is going to be a lot harder in the second phase, because that element of 'coercion', in inverted commas, of the compulsory reading list is no longer there.*
>
> (2-2-2)

Assessment

Information on how the *A Day's Wait* test had been scored and graded came easily. Mark was able to specify the required response to each of the items and the ways in which the response had been scored and finally graded.

Mark indicated any grammatical mistakes or errors on the test sheets, but they were not scored and did not directly affect the final marks. In Mark's perception, most grammar school learners are able to express themselves properly in English. That is why he feels his learners should be able to do the test in English. However, Mark again expresses his regrets of the increasing use of Dutch in course materials and tests, which unnecessarily decreases English input. (2-2-9).

Each of the ten questions scores a maximum of 10 points. If elements are missing in a learner's response, points will be subtracted from the maximum score. Item 10 (personal valuation) is generally awarded the maximum score of 10 points.

Mark discussed the ten test items in a methodical way, which allows for an ordered representation of his response to our questions. We will first present the test items, and then go into what Mark perceives as the required response.

Mark's specifications make clear that he does not easily allow for interpretations of test items that deviate from what he has in mind. We will come across some of these in the overview below.

Item 1 What is the story about? Give a brief <u>outline</u>.

According the Mark, the ideal summary should consist of the following five elements (2-2-9):

- A boy of nine suffers from flu;
- The boy overhears his doctor saying his body temperature was 102 degrees;
- The boy recalls that at his school in France he was once told that people die when their body temperature exceeds 44 degrees;

- Only at the end of the day the boy's father comes to realise that his son mistakenly thought he was going to die;
- The father tells his son about the different temperature scales of Fahrenheit and Celsius.

Mark does not warm up to a mathematical approach of assessment, such as subtracting two points if one element of the five that are required is missing. Neither it is possible for the learners to earn bonus points when parts have been phrased very well. (2-2-9).

Item 2 What is the <u>theme</u> of the story?

Mark has taught and trained his learners to be concise when mentioning the theme and use as little words as possible. *Fear of death* or the *overwhelming fear of death* score 10 points. Learners who simply mention *death* are awarded 8 points. A combination such as *misunderstanding about death* would also have received a high score (2-2-10).

Item 3 Describe the boy's <u>character</u>; what do you know about the boy?

Mark wishes his learners to be quite elaborate here. Ideal response contains a mix of plot details, such as boy aged 9, having been to school in France, and personality characteristics, such as the boy wishing to behave like an adult and planning to die like a man, only trusts his father, does not want to be surprised by death. Mark also feels it is important that the learners mention the "enormous relief" the boy feels when he learns he had misinterpreted the Fahrenheit and Celsius scales.

> *Well there are a number of elements that you can find in the story. That he went to school in France, that he's only nine years old, that's important. That he's trying to be brave, that's very important. Questions like this are important, because the learners must really justify their opinions, with a lot of words. Yes, yes. He won't let anyone in his room, will he. Just his dad, because he trusts him. The boy attempts to face death like a man. He lies there motionless, because he doesn't want to fall asleep. He doesn't want to be surprised by death. He doesn't protest, doesn't cry, nothing of the kind. But he only hears half of the well-meant stories, or the stories his good-natured father tries to read to him. And then, what also has to be included of course, is that enormous sense of relief when his dad tells him that he's only been scaring himself [yes]. And then he becomes a normal, although, what is normal, nine year-old boy again.*
> (2-2-11)

Mark admits that the maximum score of 10 points would be hard to get, but that in between 7 or 9 points were often awarded (2-2-12). He did not indicate precisely how he has scored answers that contained elements he thought were less important.

In the segment above Mark claims that the test takers have to justify why they are of opinion that certain personality traits are characteristic of the boy. The test item, however, does not ask after such a justification. In the same vein the 'enormous relief' the boy feels is not a personality trait. Besides the boy's relief seems pretty gradual (lines 104-105).

In view of the ideal response expected by Mark, one can be critical of the way in which the test item was phrased. A more explicit phrasing might have helped the test takers a little more, such as: *What have we learned about the boy? Use at least four details from the short story that refer to the boy's character.*

Item 4 Father-son relationship. What can you say about it?

Mark starts by saying that it is also difficult to get the full 10 points for this item. The response he expects here is that the learners indicate that the relationship was

very good indeed. The relationship may indeed have been good from the father's perspective. However, some of the learners had written down that the father had not observed his son very well and that he should have acknowledged his son's mortal fear. Such a point on the part of these learners might even be supported by the fact that the father even decides to go out for some good hunting Hemingway style, during which he metaphorically shot two quail and missed five in the end (l.60). The learners who had mentioned the relationship was not ideal may at least have had a point if Mark would have taken the boy's perspective into consideration. Mark discards this point of view in a rather harsh black-and-white way.

> *That [i.e difficult to get the maximum score] also goes for question four. Some kids combined those questions, which is rather logical. What can you say about the relationship between father and son? Well, it's a very good one, although some of them didn't see that, because they felt well, the father was in fact blind to his son. They had misunderstood the story, of course. And then that's the kind of moment where, during an oral exam, you can step in. Yes, because then they're completely on the wrong track. Because what they should really say there is that the relationship is a very good one. And how can you tell? Well, because the father is always very concerned, tells him to go to bed, you know. He calls the doctor and takes his temperature, makes sure that the medicine is ready, reads to him, and what should definitely be mentioned is that he, the boy that is, only trusts his father and won't let anyone else in his room, because he doesn't want them to get what he has. They could also mention that in question three, you know. The feeling that he has to be a man [Yes, almost like he's a grown-up], yes, yes [in that] yes. So that is really//. And then finally, that his father takes his question seriously. Of course at first he's totally bewildered, isn't he? Repeats his question, but then explains matters in a more serious vein, and by doing so, taking his son seriously. And then in the end the son believes him.*
> (2-2-12)

There are no doubts whatsoever that the father loves his son and appears to be very caring. From the father's perspective, Mark seems absolutely right. However, some of the learners may tend to consider the relationship from the boy's perspective instead of from the father's. In literary interpretation it is often difficult or even impossible to say who is absolutely right or absolutely wrong. Our reason for challenging Mark's plausible interpretation a little here is that we would never go as far as claiming that learners who point at the ambivalence of the father-son relationship 'had misunderstood the story, of course' (2-2-12, above). Mark does not go easy on learners who, in his view, completely miss the point.

Nevertheless, Mark says he welcomes literary interpretations as long as they are justified by the learners. According to Mark *A Day's Wait* does not allow for a lot of variations in interpretation.

> *Yes, well, there is, maybe not so much with this story, but definitely with novels, you have a certain way of viewing a certain person, a character. If they can support their interpretations, it's fine with me. But of course, you do have quite an advantage over them, especially when you have come to know a story very well, so that you can talk back to people and tell them: 'Well, why don't you take a good look. Don't you think that's a bit far-fetched?' or 'You've just been making this up as you went along'.*
> (2-2-22)

Item 5 What is the crucial line of the story? Why?

Here the justification question that was missing in item 3 is raised. Mark is explicit about what the crucial line of the story is.

> *If you got that right, the crucial line, you would usually come up with an explanation as well. And of course that's in line 87, isn't it? 'About what time do you think I'm*

> *going to die?' And then everything is revealed at once. And the father can't believe his ears: (In English) : 'What? "About how long will it be before I die?' And yes, then everything becomes clear to the father, but also to the reader, ah, is it like this? And yes, so that's a question where it's really all or nothing. Because if you choose another line, well, then you're just plain wrong. When I made this test, I remember, I did wonder, well, isn't there any other line possible and then I found this one: (Mark points at ll. 41-42: 'You don't have to stay in here with me, Papa, if it bothers you.'), Rereading that line, it appears to be very important as well. And then I thought, well, to prevent that, then I'll just turn that into a separate question, you know, because both those lines to me are crucial lines. But I still find this one, line 87, the most important. Some people wrote down line 89 (i.e. "How long will it be before I die?"). They came close. But then you're a few seconds later already, whereas here, yes, just like the father, the reader can hardly believe it. So I subtract a point from that, or maybe two points. But, they're still on the right track.*
> (2-2-13)

Mark does not see the added 'why' as a separate question. It is seen as an item that derives logically from indicating the one and only crucial line. Mark does not value any other lines apart from lines 87 and 89. In order to ensure the learners do not come up with "You don't have to stay in here with me, if it bothers you", Mark has used the line in item 6.

Item 6 "You don't have to stay in here with me, Papa, if it bothers you." (ll 41-2) What is <u>it</u>?

Mark does not accept as a response "sitting here with me", but wants the learners to add "watching me die." (2-2-14). Therefore, Mark is mainly interested in what the word "it" refers to from the boy's perspective. Therefore, a question such as "What does 'it' refer to from the boy's perspective?" would have added to the validity of the test item 6.

Mark says that learners who are on the right track easily score 7, 8 or 9 points for "comprehension items" 5, 6 and 7.

Item 7 Explain the last line of the story; "…he cried very easily at little things that were of no importance."

Mark wishes his learners to interpret the line as "that enormous relief, the letting go of the role of heroism he had assigned to himself", and become a child again.

> *Yes, he is so tremendously relieved. (In English) 'He is relieved, he, he can let go'. He becomes that little boy again and definitely because he has been so forcedly stoic, so manly. Yes, that has to be expressed somehow. [yes] But then you also see children who just find it hard to identify, and then it's striking, well, I don' t know if you're even allowed to say it, but that girls are better at that than boys.*
> (2-2-14)

In order to arrive at this answer, Mark feels the learners have to be able empathise and identify with a character. In his opinion, female learners are better at this than male learners. According to Mark, dealing with literary texts at secondary school helps learners explore what they think and feel. These are the moments when he sees himself as an educator (Du: pedagoog) rather than an instructor (Du:docent).

Mark offers examples of learners exploring racism and discrimination, with himself in the role of the educator. This is a role he values so much more than just being an instructor or classroom supervisor. He feels literature helps people explore moral dilemmas

> *And you have to make sure the learners can relate to that, don't you? Say, OK, gypsies? No problem. Refugees? No problem. But now they are moving in as your*

> *next-door neighbours, right. That happens. There are refugees. More bikes have been stolen. And people start pointing their fingers, you know. But of course you don't discriminate, do you? You don't do that, do you? Yes, or do we in fact - ? Because it's very easy to be for something or against it in theory, but actions speak louder than words [Clearly]. And these are the things that, well, then you're more of an educator than an instructor of English. We mentioned that last time. Where you get to mould learners in a certain phase of their lives in ways that you think will really be of value to them. And one learner picks up one thing, and another picks up something else. But that's disappearing more and more these days, and separate subjects are supposed to be covering that. And I really find that a pity sometimes, because that is one of the most attractive and motivating aspects of being a teacher.*
> (2-2-15)

Mark once more expresses his regret of the reduction of his role as a teacher from being an educator to being a mere instructor.

Item 8 <u>Construction</u> of the story. The story can be divided into <u>three</u> parts.

a) When does part two begin? When does part three begin? Mention the numbers of the lines;
b) What is the function of part 2?

Mark scores 5 points for the a-question and 5 points for the b-question. Item 8 concerns the second part, which starts at line 47 and ends with line 62. In this part the father goes out hunting on this bright and cold day. The icy surface is so slippery that he gets knocked off his feet twice. The father hunts and kills for pleasure. Death does not seem to affect him. The predominant feeling when he returns to the house is happiness. Having missed five quail he knows there are still some animals left to shoot on another day.

According to Mark, few learners answered the b-question correctly. The correct response for Mark here is that the learners say that "the father is also concerned with death and rejoices over the quail that are still left there to shoot on another day."(2-2-16). Mark appreciated that a few learners mentioned the irony of the scene or its sympathetic landscape, i.e. the icy and cold day.

Item 9 Explain the <u>title</u>.

Mark has scored the maximum of 10 points if learners have indicated "waiting for a long day for the seemingly inevitable to come." (2-2-17). One learner had indicated "time killing" as a function of the hunting scene. Mark appreciated the pun with 1 point.

Item 10 Do you think this is a good story? Why (not)? <u>Personal valuation</u>.

Mark says that the response to this question is personal. If arguments or some justification are given, the response almost always scores 10 points.

Assessing how adolescent learners interpret *A Day's Wait* is likely to be affected by how the assessor interprets the story. Mark was explicit and straightforward with what he saw as the correct response. Nevertheless, there were some occasions on which we took up Mark's interpretations. In chapter 11, we will discuss the essential measurement qualities of reliability and validity in relation to the findings of this study.

Mark was detailed on the way in which he had assessed the literary test. Next, we will provide some details on how the test was evaluated.

Evaluation

Mark indicates that there has been no time to discuss the test results with his class. He blames the second phase innovations for this. A classroom evaluation of the test would take valuable time off the preparation for a next *Unicom Finals* test on

idiom, grammar and reading comprehension. These tests are administered every four or five weeks. Mark says the following in a tone of apology and regret.

> *It's a shame, and some of the learners remarked on it, that this new system, yes, or you just have to go ahead and do it of course, but there is no time to discuss the story in class. And yes, I find that a real pity. You're, yes, you are more limited, you know. And yes, perhaps I should come up with a solution some day. It's the first year, because then you take valuable time off of preparing for one of those Unicom Finals' tests.*
> (2-1-19)

Mark is well aware that some of his learners would like him to discuss the test and its results in detail. He is also aware of the fact that he does have a choice as a teacher. It is a dilemma which was not solved in the year of data collection. Tests were assessed, but rarely ever discussed and evaluated afterwards with the learners and by the learners themselves.

Mark tells me he does not suppose learners to do a lot with the test once it has been returned. Besides the happy few who might remember Hemingway, like the story, and realise that they have learned how to read, he does not count on any positive washback effects for the majority of his learners. Mark simply accepts that learners are calculating beings, primarily interested in the required knowledge and skills in view of future tests to come.

> *I: What do you expect learners to do with the test once it has been returned? ..Or what do you hope they would do?*
>
> *M: Well, I don't expect them to discover Hemingway or whatever. No, I think that the value is mainly, for those learners concerned, that they learn how to read and that they might say, hey, that was a good story. And they might remember Hemingway. But other than that, learners are quite calculating most of the time. They are concerned about the marks, and do not share the teacher's goals. Like, hey, you got a, say, 7.4. So, for a first impression of literature and the required skills, you've done quite well. And maybe that will help you when you receive your first mark on a novel, you know. But other than that? No, I don't think they will do much with their returned test.*
> (2-1-25)

Mark thinks that marks are important to almost all of his learners. He feels that positive marks on the *A Day's Wait* test, might be some indication for learners how they are likely to do on tests that concern full novels.

Evaluation of the short story test was already discussed in terms of possible washback effects that Mark perceived. In the final two sections on the test, we will have a closer look at how Mark interprets test washback of the *A Day's Wait* test.

9.4.4 Washback on CLE

Mark puts forward that the skills of listening, speaking and writing were integrated well in the lessons that preceded the test. All of his literature lessons are in English and the learners respond in English. Therefore the primary focus is on listening and speaking skills.

Writing skills are practised because learners have to produce summaries and/or take notes when short stories are discussed in class.

> *I: How communicative does a learner become after all the preparation and the test?*

M: In the preparatory phase, which is an essential part of teaching literature as well, and I have said so before, speaking is very important of course, and so is listening.

I: Your lessons are all in English? Literature is discussed in English?

M: Yes, yes. Well, the literature lessons are in English. And writing is practised when I ask them to write summaries. Some children also write about a character, or about construction and so on with the first stories. It's mostly oral, but if you want to assess something, yes, then it's mainly written. Except for the oral exam. So all the skills are there and, yes, it's so important even, I have to mention this, that this school mentions in its school profile that literature is of central importance at this school and that we practise the language skills by way of literary tests. And that means that in forms five and six, two out of three English lessons are filled with literature and the other lesson focuses specifically on listening comprehension, or reading comprehension, or …

I: The skills are applied to literary texts//.

M: YES, YES, YES (loud), so there's a lot of literature.
(2-1-28)

Mark does not in any way refer to the fact that the learners have to put across their ideas on *A Day's Wait* in English in such a way that he understands what they mean.

As is the case with Joy, the actual test itself is not seen as affecting the learners' communicative skills in English. Like Joy, Mark feels that the lessons that preceded the test educate learners how to listen, speak, read and write in English. The actual test is not seen as a factor in this process.

9.4.5 Washback on LA

Mark had already expressed the need for an educator when learners learn to read and appreciate literature. We have already learned that Mark believes in teacher demonstration and exemplification when learners learn. It requires an active role of the teacher by way of direct teaching. He particularly thinks this is important when learners have to be engaged with a certain short story or literary task. Mark feels you cannot simply replace the teacher's role by written instructions.

And yes, what you see more now, because they work independently, so they can manage fragments or short stories, with questions they have to answer themselves. But that doesn't come alive then. You have to read a story like that OUT LOUD (emphasized). And you need time to do that. But anyway.
(2-1-24)

Mark does not feel the learners have to be autonomous to prepare for the *A Day's Wait* test. Learners are well-prepared if they actively participated in the lessons, in which four or five short stories were dealt with. He expects some of the learners to perhaps have studied the literary terms. This is how Mark reacted to our question in how far the test affects how learners learn on their own.

M: Yes, what do they have to learn from this? Nothing, really. I think that a few learners might have gone over the literary terms again at home. But that's what I advised them to do. The actual practicing happened in class, by participating in the stories that were dealt with. So what does it add to working autonomously? That mainly occurs in the higher forms, where you sometimes ask them to write a reading report or an essay about a book or a story or about poetry and then, well, they add to

the knowledge that they have developed over the past years. And, yes, they become more autonomous.
They do projects, together, or alone. We just did a war project and there are some really good products from that. But yes, this test is really part of the build-up of literary knowledge and skills. But how we'll continue next year I don't know yet, because we, well, because we still have to develop the new programme for assessment and summative evaluation.
(2-1-21)

The focus on short stories is intended to help the learners prepare for tasks that require more independence and self-regulation in the two years to come. That is why Mark has deliberately directed and controlled his learners when they prepared for the test.

M: …in preparing for this test, there was a lot of teacher direction, because they have to be taught the specific knowledge and skills. Then in the fifth and sixth forms, they are allowed to work more in pairs or in groups. And also in the fourth form they conclude the year in groups of four.

I: The project you already referred to, didn't you? [yes] It will be groups of four that you//

M: Yes, I think//. It depends on the number of kids. If it's 28, it'll be 4 x 7 of course. But three is possible as well.
(2-1-30)

Interestingly, Mark's fourth-form learners were to be allowed more autonomy and self-regulation at the end of the year, when they worked in groups of three and four on a novel of their preference. This particular test had been selected by Mark as the third test for discussion.

Mark admits he knows very little about *how* and in *what ways* learners learn from mistakes or errors in the course of teaching practice. The only concrete example he mentions is the skill of making summaries. This is among the 'learnables', and he expects learners to explicitly learn from the mistakes they make when they attempt to write effective summaries.

I: Do you yourself as a teacher have insight into how learners learn from mistakes they make when they practise with the other short stories you have dealt with?

M: Not really.

I: So this is in fact the first time that you yourself actually see what//

M: Yes (loud), they are// maybe writing the outline is the clearest and most concrete example, really. You can really, not that I look at their exercise books, or whatever, but learners can actually see for themselves, hey, the first time I had, say, ten lines and I needed fewer lines to summarise the story. I think that maybe half the class was asked to read their summaries out loud. But other than that, it's really, well, a lot of practice.

I: And making sure that you get it yourself eventually? [yes, yes[What the essence really is?

M: Yes, and intuition has a lot to do with that.
(2-1-22)

Besides information on how well the learners are able to summarise short stories, Mark says there is hardly any formative assessment and evaluation when the learners are prepared for the short story test. Practice makes perfect, seems to be the adage.

At the end of the interview Mark is once more explicit about the kind of autonomy he is after with his initially teacher-directed approach.

> *I: Yes, well the last question is about self-regulated learning. You've already said quite a bit about the subject. The degree of autonomy. In how far does this test encourage the learner to regulate their own learning?*
>
> *M: Yes, later you will have to read some books on your own and you'll have to do that wholly independently. And, yes, hopefully it encourages learners to read. And that's another thing that you'll have to do independently and afterwards and meanwhile, I hope you won't just say yes, finished this book, next book, but that you'll try to see, you know, preferably with others.*
>
> *I: They should be affected by what they read. [YES (loud)] They should be able to relate these feelings to their own experiences and development.*
>
> *M: YES (loud), exactly, yes. But that's something that you, yes, sow (with emphasis) and it doesn't mature until after you leave this school.*
> (2-1-31)

9.5 Test 2: The *Unicom Finals* unit 8 test

The second test Mark had selected for discussion was a regular test based on the course materials, which tested for idiom, grammar and reading comprehension. It is typical of the tests his fourth-form grammar school learners get every four or five weeks. The learners work their way through the texts, tasks and activities of one or two units of *Unicom Finals* largely on their own. The *Unicom* course materials are generally known for a solid but traditional approach to English language teaching, with a firm focus on grammar and idioms. The materials have recently been revised to fit in with the curricular and didactic innovations of upper secondary education. The revised edition of *Unicom Finals* was going to be used for the first time at Mark's grammar school in the year of data collection.

Mark expects the materials to be helpful in meeting the new requirements of upper secondary education. That is also why he and his colleagues welcomed the revised version of the course materials and the tests that came with them.

9.5.1 Justification

Unlike Joy and Pete, who had constructed all of the tests themselves, Mark had worked with test items that came with the *Unicom Finals* package. His justification for presenting this test therefore focuses on why he had decided to use the tests that came with the course materials.

> *M: Well, this year we started with new course materials, focused on the Second Phase. This was the case for more school subjects. Because of lack of time, you often don't get down to review course materials in any depth before actually using them in everyday practice. Besides, you can't really assess course materials correctly until you've worked with them for at least a year, or even years. And, by extension, you can say the same about tests. The tests come with the course materials.*
>
> *I: It's working (interviewer checks whether the minidisk is recording)*

M: It's working, it's being registered. They come with the package, so you actually rely a little on the expertise of the persons who constructed the tests. And you kind of depend on that, because the preparation, the reflection, has been done by experts, who are much more aware of the exact contents. It's also a matter of lack of time, otherwise you could make the tests yourself, as we partly did at the beginning, but eventually we decided to just use the tests that came with them. Two versions of tests are provided, A and B, and some of the assignments are too elaborate to be administered in one lesson. So we make a selection and in doing so we use selection criteria which we have developed over the years, which are the criteria of knowledge, skills and reading comprehension.

(2-2-2)

Over the years, Mark and his colleagues allegedly decided that a test should be selected on criteria of 'knowledge, skills and reading comprehension'. Apparently, Mark does not see reading comprehension as a skill, which give the selection criteria a slightly rhetorical ring. Mark's selection criteria seem to have been primarily pragmatic. He seems to have selected appropriate idiom and grammar items from the test, to which a reading comprehension text and test has been added. Mark had also handed in other Unicom Finals tests he had been using in the course of the year. They had all been structured in the same way.

Mark spontaneously compares the revised edition of *Unicom Finals* to the edition he had been using up the year before. He concludes that the parts meant to be learned by heart had decreased. There is a less prominent role for vocabulary acquisition and it generally includes more skills. Mark refers to these skills as follows: 'The skills involve creative involvement with English by way of writing a letter or translating something.' (2-2-3).

The test package that comes with the course materials also includes reading comprehension tests. Mark sees these tests as excellent preparations for the havo final examination reading comprehension test the learners have to do as a school test at the end of their fourth year (2-2-3). Mark is of opinion that the levels of the listening and reading comprehension tests from *Unicom Finals* are about as high as the levels of the tests the learners generally get in their sixth year. Mark does no longer see an increase in level of difficulty of the texts offered, which he felt was characteristic of previous editions of *Unicom*.

According to Mark *Unicom Finals* has been structured in such a way that the odd chapters introduce new grammar points and that the even chapters focus on consolidating and internalizing the grammar that had been dealt with before. This means that learners generally work their way through the even chapters on their own, with little coaching or guidance by the teacher.

Thus, for Mark the course materials to a great extent determine what is to be taught, at what particular time and in what particular way, which he feels as convenient in this year of change.

9.5.2 The knowledge, skills and insights measured by the test

The test consists of three parts. The A-part concentrates on grammar and idiom, with a strong focus on translation from Dutch into English. The B-part is a letter in Dutch filled with vocabulary the learners have studied in unit 8. The letter has to be translated into correct English. The C-part is a reading comprehension test in which the learners have to fill in twelve gaps in a lengthy text. For each of the gaps they have to select one of the two options offered.

This is Mark's initial response to my question what is measured by the test in parts A,B and C.

M: Well, they're the things that I have already mentioned. Knowledge, rote learning, but also grammatical understanding. Tenses and such. All the basic things they need.

I: Let me write that down, so that I can keep up.

M: It's also idiom, which we agreed that we would begin with at the beginning of the fourth form. In the back of the book you have these 'useful phrases'. To some extent you have to feel your way, isn't this too much, or, well, we want that to be a part of the test from the beginning of the year. Well, these are some grammar exercises. You shouldn't forget that the course materials have been constructed in groups of two, you know. Sets of chapters that belong together, where every even chapter repeats what has been presented before, so that these chapters can be worked through independently by the learners, that is to say the preparation of it. And therefore this part is rather brief. Because chapter 7 was tested in the same manner and so you could find more grammar in it.

I: The test on unit eight has a certain repetitive effect, which allows you to point out, well, to someone who has made a lot of mistakes in this part? [Mark nods in agreement] I see. They just hadn't studied [Yes, yes, yes] that item, or hadn't retained it. Yes. Does the same go for part B, because for the most part these sentences can literally be found in the 'useful phrases' in the back?

M: Yes, yes, that's right. So hard work really pays off there. And these are also exercises that we, I think we got some from the tests that came with the text book, or they came straight from the exercises in the workbook and uh.
(2-2-12)

Mark feels the grammar points and vocabulary tested in parts A and B can be reproduced if the learners have studied well. He sees the grammar and idioms as reproductive knowledge elements.

Next, we will outline how Mark views the required knowledge and skills tested in the three parts of the test in more detail.

Part A: five vocabulary and grammar/translation exercises

Part A consists of five test tasks. The test items of assignment I concern the reproduction of the main forms of six irregular verbs, i.e. infinitive, simple past and past participle. The three forms are elicited by presenting six Dutch infinitives that the learners have to translate.

Test task II is a translation exercise from Dutch into English, in which the learners have to reproduce 'useful phrases' and some discrete idioms meant to have been learned by heart.

Task III is also a translation exercise focusing on vocabulary. Now the learners have to translate English sentences into Dutch. I ask Mark whether the relatively brief sentences of task III had been there with some purpose.

M: Well, no, not really. Often English sentences are in fact shorter than Dutch sentences, you know. In any case. But no, I think that's just a coincidence. It's a coincidence.

I: Yes. There is no demonstrable difference if you're talking about learning for 2 and 3. Exercises 2 and 3.

M: No, because they're all from the same number of sentences that they were supposed to study, both Dutch/English and English/Dutch, so no. It's just a coincidence that there are some shorter sentences here, yes.

I: That the sentences in 3 are somewhat shorter. Yes, yes. You know, I don't know whether it's true or not, but learners are suspected to do better receptively than they do productively.

M: Yes. This really forces them to study very accurately, doesn't it.

I: Yes, definitely.

M: And we combined a number of things here. There were some shorter sentences in the book, I believe, but we have two sentences: 'gefeliciteerd met je fantastische prestatie' (congratulations on your remarkable achievement) was one of the sentences and then we had 'ook namens de rest van de family' (also on behalf of the rest of the family), so that was listed as a separate sentence in the book, or as another sentence. 'Ik feliciteer je namens de rest van de familie' (I congratulate you on behalf of the rest of the family), so we just combined the two.

(2-2-15)

On second thoughts, Mark remembers that he had combined two of the English phrases that the learners had studied into one Dutch phrase. Translation of Dutch into English is characteristic of the test.

In the first interview Mark had already indicated that he was a teacher at a *grammar* school. To Mark this literally means that learners have to be conscious of grammar and grammatical similarities and differences between languages. The learners are used to translating texts, both in their lessons of Latin and Greek and in their English lessons. The translation sentences and tasks in part A and B indicate that Mark aims at high levels of grammatical and lexical knowledge. Mark feels these high-level tasks create a knowledge base that may help learners to become better and more creative communicators in English. Mark is well aware that the actual transfer of such knowledge to spontaneous language production is far from self-evident. He also realises that learners study a lot of idioms that are not necessarily directly relevant to the English they are using.

I: In our first interview you talked about a grammatical and idiomatic foundation, didn't you, to which additional knowledge is added. Do you feel that is the case with the useful phrases, idioms and the grammar, which have to be learned before the learners are ready to move on to freer assignments?

M: Yes, when you offer idiom, and useful phrases and such, you always have to consider the following question, how much of it is actually 'usable', you know? "Thank you teacher", you know that song. "The millions of words that I will never use". That's the sigh heaved by the learner who has to know all those phrases by heart, and who wonders: 'Is this really modern English and will I ever actually use it? And of course that learner has a point there. But on the other hand, you can't escape the need for a practical command of phrases and idioms to be able to communicate. So that will remain a constant battle of course.

(2-2-23)

A little later in the interview, a teaching experience is shared with Mark that learners who have memorized 'useful phrases' at regular stages in their learning, start using these phrases easily and spontaneously when they speak or write. They seem to do so 'on the spur of the moment'(3-2-25), without any conscious planning. Mark acknowledges this particular experience of ours a number of times, and responds:

M: Yes, that's true.

I: And it seems that, well, knowledge seems to play a greater part than uhm.

M: Yes, but it also has to do with a step in the process that you can't really test, that is, of course you can acquire knowledge, but you have to reach one step beyond. You have to internalise it, so that it really pops out automatically when you want to use it. And there are learners of the really industrious, diligent, reproductive kind who have a phenomenal memory, but two days later they've lost what they have memorised. Learners like this just peak at a certain point in time. Yes, then it seems like they've got this amazing vocabulary, but what you're really awarding with that 9 or 10 is their short-term memory. And the learners who are less intensely trying to soak up everything like a sponge, and who look at things more with their brains and don't try to translate everything literally, but appropriately nevertheless, you can really notice that knowledge has sunk in a lot more with these learners. And they have actually mastered a lot, perhaps without even being conscious of it.

(2-2-25)

Mark is aware of the problem of transfer. Just as Joy and Pete, he aims at transfer of knowledge to real communicative tasks. Mark feels that he cannot really test or assess the ways in which learners have internalised knowledge. Mark is also well aware of the limitations of mere reproduction of knowledge that favours learners with excellent short-term memories. Many of these learners do not succeed in retaining what they once learned. Their learning efforts do not automatically result into the expected levels of proficiency. On the other hand, learners who learn more consciously and less reproductively are better learners according to Mark. They manage to retain their knowledge for a longer period of time and are able to use this knowledge creatively.

We will now continue with Mark's comments on test assignments IV and V. The first grammar point tested was knowledge of the three main forms of the irregular verbs. A second grammar point is the use of adjectives and adverbs, which is tested in tasks IV and V. Learners should do more than simply rely on their feelings and intuitions why one particular form or structure is to be preferred over another. Mark feels his learners should be able to reason which particular structure has to be used in a given test item. In order to reason with success, the learners have to know to what antecedents adverbs and adjectives generally refer to (e.g. adverbs →verbs, adjectives, pronouns, or other adverbs; adjectives→ nouns, use after the copula 'to be' and verbs of perception such as *smell, appear,* and *sound* that can all be replaced by the copula 'to be' for the person who perceives). The issue here is the ability to reason on the basis of a systematic rule taught and/or studied or rely on principles such as automaticity or 'gut feeling'. We will address these issues in more detail in our discussion chapter. As we have seen in chapter 3, the ability to reason is a parameter of learner autonomy and may be relevant to washback as well.

Task V assesses a third grammar point, i.e. the translations of Dutch 'De betrokkenen' (Those concerned) and 'De aanwezigen' (The people present). Besides, item number 4 includes the tricky translation of Dutch "Wij kunnen onmogelijk" into "We cannot possibly", in which the negation is added to the auxiliary.

The next part of the test claims to focus on writing skills.

Part B: Letter writing

Part B does not appear to promise what the heading 'letter writing' suggests. It is again a translation task, in which a Dutch letter has to be translated into English. The main difference with test task V is that the sentences are now related to one another.

As far as the knowledge and skills are concerned, the text presents a mix of the vocabulary and the grammar that has been studied. Yet, quite a few of the grammar points included in the letter are different from the grammar points studied in the unit, which had been tested in tasks IV and V. Besides elements of vocabulary and spelling, some of the grammar points that stand out are:

- correct word order
- passive constructions, such as “gebruikt werden”(were used), “worden gefokt”(are (being) bred) or “worden zelfs afgemaakt (are even put down)
- absence or presence of the indefinite article: “Wat een verachtelijke mensen” (What despicable people) “Ik word lid”(I become a member)
- Tags: “Vreselijk, hè?”(Terrible, isn’t it?)
- “niet durven gaan”(daren’t go, do not dare to go, are afraid to go, fear to go)

Mark said the letter writing assignment appeared to be the most difficult part of the test, which may in part be caused by the additional grammatical pitfalls mentioned above.

The third part of the test assessed the learners’ reading comprehension.

Part C: Reading comprehension

The reading comprehension test is a gap-fill in which the learners have to choose between two given options. Mark mentions three criteria that help him select an effective reading comprehension text and test (2-2-27):

- The learners should have enough time to do the reading comprehension tests, which always come as the last tasks in *Unicom Finals* tests;
- The test format should resemble the formats of test items of the CITO reading comprehension examinations the learners have to sit at the end of the 6th form;
- The tests the learners do should gradually increase in level of difficulty.

Mark feels the latter criterion is the most difficult to realise. He feels it is hard to assess the level of a test before the actual test is administered. Once it has been administered and scored, an indication of the level of difficulty becomes easier. The way in which Mark scores and grades is norm-referenced rather than criterion-referenced.

Mark defines the knowledge and skills that underlie reading comprehension as follows:

> *Yes, I think it’s reading comprehension and that’s an easy answer, because what exactly is reading comprehension, irrespective of how you assess it? And this may even be one of the easier formats, you know, with an A or B option. I think reading comprehension is only in part related to the foreign language. It is also dependent on the subject matter of a text. In addition, it also has to do with general intelligence, if you will. The ability to combine, common sense, well, you name it. Concentration plays a part, too. They are aspects you do not mean to assess, but which are inescapably linked to this way of seeing what learners can do. The same is true of listening. If you don’t know a thing about the subject, then you’re way behind already.* (2-2-35)

Mark feels that reading comprehension is made up of four components. A first component is the foreign language knowledge and skills the learner requires to

understand a text. A second component is the learner's topical knowledge. The more a learner already knows of the subject matter, the easier (s)he will understand the foreign language text. Third, Mark refers to general intelligence and the ability to infer meaning from the text by combining the separate elements that carry meaning in a text. As a final component Mark mentions concentration. Learners who are able to concentrate on a text for a longer period of time, stand a better chance to develop their reading skills.

9.5.3 Construction and use of the *Unicom Finals unit 8* test

We will again address six aspects related to the construction and use of the regular course materials test selected by Mark. Most of the sections on construction and use will be relatively brief, because information has already been presented in the quotes we have dealt with so far. The exception is the section on evaluation. It is not so much spent on how the actual test was evaluated, but concentrates on how Mark evaluates his role as a teacher in the second phase.

Construction

The test was primarily based on the unit 8 test that came with the course material package of *Unicom Finals.* The test was slightly adapted. First, it was ensured that the test could be administered in a 50-minute lesson. Second, some Dutch translation sentences or phrases were added to existing test items.

Expertise

Because the test was not constructed by Mark himself, it required relatively little expertise. The slight alterations were made by Mark and/or the colleague he was cooperating with in the fourth forms. Mark puts down his testing expertise to practical experience, which has enabled him to routinely make any changes.

Conditions

The *Unit 8* test was a paper-and-pencil test administered in the regular test week consisting of five days. The learners were given 60 minutes to do the A, B and C-parts of the test. The use of extra materials such as dictionaries had not been allowed.

Preparation

We asked Mark how the learners had been prepared to the test. Mark objected to my using a passive construction here and told me the learners themselves had very much been working on their own.

> *Yes, well I wouldn't use a passive construction here. The second phase is all about learners preparing themselves and them being offered the support of extra materials if they need it, you know. There's a CD, they have the answer sheets, they have the explanation as to why a certain answer is correct and then finally you also have this guy who sits there marking tests or, I don't know, keeping the peace. So of course there's also a teacher available to them, and yes, they do all make use of that. But they prepare for the test on their own. And when there is something to be done together in class, then, well, then they're more or less forced to participate. But that's not, I couldn't tell you the frequency, but that's not very often. And that isn't what the book aims for, you know. Like I said, Unit 8 is something they should really be able to do on their own.*
>
> (2-2-39)

Mark is far from pleased with the reduction of his role as an educator preparing the learners adequately to tests. He elaborates on five restrictive factors: loss of direction and control over the learning process because the learners largely work on their own, strict adherence to the Unicom Finals course materials, the programme for

assessment and summative evaluation he was forced to draw up for his fourth forms, the limited use the learners could make of the computer room and media center, and finally the loss of literature as the primary focus in his upper secondary classes. We will present more details on Mark's criticisms in the section on evaluation below.

However, Mark's learners seem to have been familiar with working on their own before the introduction of the second phase. When we discussed part C, the reading comprehension test, Mark mentioned how he is used to teaching his learners to develop their reading comprehension. Mark plenary discusses some reading comprehension tests in class. Then the learners are on their own. He repeatedly tells his learners that the usefulness of foreign language practice is dependent on the concentration and effort they put into the reading comprehension tests they practise with. This is what Mark's learners have been told from the second form onwards. It is up to the learners to show that practice makes perfect.

> *Yes, then ultimately you come back to the same thing. In a certain sense they're with it. Or maybe they're not. We run through a couple of texts together. We read those together. We talk about them, but for the most part they practise on their own. And you can keep repeating, which by the way is something you start doing in the second form, that some exercises are only useful because of the way in which you do them. Once you have finished them, you correct them, yes, and that's it. You can't learn it all over again. You have to do them. The training, the value of the training is inherent to the training. The value is not in videotaping the training and saying this and that afterwards. No, you just have to do it a lot. Practice makes perfect. And the text book offers a lot of practice material, but at the same time you could say, yes, but when learners look at that, and very often they're long assignments, such as the ones in this test you know, it's a big exercise and they go 'pffff', and don't really feel like that right now.*
> (2-2-28)

Mark says that the value of training is in the actual training itself. He does not seem to advocate any formative reviews of the learning process on the basis of the tasks or activities a learner has carried out: 'Once you have finished them, you correct them, yes, and that's it'. Mark's view of assessment is predominantly summative. At the same time Mark acknowledges, that learners are likely to be discouraged by some of the reading comprehension tests they have to do. This is particularly the case when exercises are long and are about subjects the learners are not in the least interested in. Mark feels it is the learners' responsibility to be 'with it'. The learners have to make sure that they tackle enough text and tests so that their reading comprehension will improve.

Mark's particular focus on learning how to read seems primarily product-oriented. The learners are given almost full autonomy here, with Mark as the task master making sure the learners can do their assignments in peace and quiet. In the discussion chapter we will address whether this type of autonomy is likely to lead to better foreign language readers, or to better foreign language test takers for that matter.

Assessment

For the A-part of the test, with its focus on idioms and grammar, 1 point was deducted from the maximum score of 10 for every 3 errors a learner has made. For the B-part, i.e. the grammar-translation letter, 1 point was deducted for every 2 errors. In the case of the reading comprehension test, 1 point was deducted for each and every error made by the learners.

Evaluation

Mark has not evaluated the test. It is up to the learners what they do with the test, once it has been returned. Mark says that he simply lacks the time to discuss tests that have been scored, graded and returned.

> *I: What do you expect the learners to do with this test after you've returned it?*
>
> *M: Je ne sais pas. They learned in the junior secondary forms that they could learn from old tests and exams. And you see that some people really do save all of them. But then, saving them or throwing them away really boil down to the same thing, don't they?*
>
> *I: Yes, that's clear.*
>
> *M: So, I don't know.*
>
> *I: You, don't know. And you don't see anything that you as a teacher could do about this?*
>
> *M: Well, I do wonder whether we shouldn't spend a bit more time on discussing the test results. But then again, logistically speaking, in the test week you get back tests from nine different classes. Checking fourth- form tests is a lot of work. It's very detailed. So it could be that you spend two weeks on checking tests of one class and one week of another, and then it's time to present their study schedules again, because the learners have to be handed a sheet with all that's expected of them for their next test week, so, yes, it's just a real hassle. To put down your work and to say, well guys, now we're going to spend some time on this, whereas it isn't even necessary for those who studied well. It's just a loss of time.*
>
> (2-2-51)

So we can be brief on how the test was evaluated by Mark. His position on test evaluation can be interpreted more clearly if one takes his criticism of the second phase innovations into consideration. Earlier in the interview, Mark had been elaborate in evaluating the second phase from his perspective as a grammar school teacher. Mark has highlighted four aspects: loss of teacher control and supervision, strict adherence to the course materials, PTA, school facilities, and the diminished time for literature.

Loss of teacher control and supervision

As an adept of direct teaching, Mark regrets it is now more difficult for him to check what exactly his learners are doing when they carry out the required tasks and activities. He feels his control over what learners do and how they can best proceed has largely disappeared.

> *M: I have to say that I don't feel very happy as a teacher in the second phase. And that I am still teaching the fifth and sixth forms with great pleasure this year, well not the sixth form this year, but the next, because I'm in control of the programme there to a much greater extent. And I can do my own thing in a positive way, whereas here I'm more or less forced, also because of the new demands and because of the course materials, which are not too bad by the way, have been banned to the sidelines more and more. And of course I can still decide to enter the field, but yes, that would be in violation of the new way of working a little.*
>
> *I: I see, which makes you feel like you don't know what learners really know and are capable of until they have shown it in a test week.*

> *M: Yes, then slowly you get to know them as learners, but mainly through the marks they have collected in your diary. And I find that, you become more of an instructor, I've mentioned this before. It's hard times for the educator these days. You're not building up relationships. You're more replaceable.*
> (2-2-29)

Mark regrets that he does not really get to know his learners as he used to. The problem is not a lack of tools to assess learner progress. The results of the summative tests in the end give Mark the information he is after. What he deplores is the lack of teacher-learner interaction. This makes it so much harder for him to really get to know his learners.

> *Yes, well, you have been given the tools, but it's mainly the process that I deplore. And then I'm speaking from my perspective. The way in which I perceive my profession, as a subject expert, as someone who needs to be in touch with his audience. And yes, the way that the second phase has organized education, has forced me to take another position in the field. And that's something that takes time for me. And the first signs are that teachers have been forced to stay near the sidelines too much. But I don't want to say anything about the learners' progress, because when you evaluate them after a test you have a fairly pretty good view of that. But it's the learners themselves that you don't really, well, you know them by name of course, but not really as individuals. That's my greatest disappointment. That's what's been lost.*
> (2-2-33)

Mark does mention one positive aspect of having the learners work and learn on their own. He says that from time to time he is pleasantly surprised by what the learners are able to do and actually show after having worked on their own. (2-2-31).

Another aspect Mark is critical of is the didactic approach of the course materials.

Strict adherence to the course materials

At the beginning of the school year, Mark and his colleagues had opted for course materials that would help learners to partly work and learn on their own. In retrospect Mark feels he has relied too heavily on these materials.

> *M: I find it difficult to accept that I don't have a clear view of what learners are really doing. Also because they're often busy outside the classroom, although sometimes I can be pleasantly surprised by what they can do and show. Yes, you really have to ask the target group themselves of course. It's quite difficult. And definitely with learners you see for the first time when they enter the fourth form. That was the case with two of the three classes that I taught this year. Yes, and then you notice that you don't build up much of a relationship with a class like that. And once again, I'm partly to blame, because I should have gone about it in a different way and I shouldn't have let myself be governed by the course materials.*
>
> *I: But as you mentioned, you start with new course materials. You can't really say anything decent about the text book when you have just been using it for a year.*
>
> *M: Yes, and of course a text book should really be a tool that you yourself are in control of. You decide what to use. And in the beginning you feel like you should just do it all. And well, this means that the text book has a huge impact on how the learners are taught, more so actually than on the way in which you teach, I would say.* (2-2-31)

Programme for assessment and summative evaluation (PTA)

Mark feels that learners have the right to be explicitly informed about the ways in which they are going to be tested. However, he does not feel it is necessary to draw up a detailed plan or programme for his fourth forms on *what* is going to be tested, at *what* particular time, by *what* particular test. Mark feels that the PTA takes away the flexibility that is required in the light of the curricular and didactic innovations.

> *M: Yes, well, that's another limiting factor. We're working on that as well, on drawing up the PTA for the coming years. And you really have to make that rather explicit, because learners have a right to know what exactly they can expect. But then you're really stuck with it, even though you haven't drawn the programme up in detail. So it really takes time and, yes, this is a year for experimenting and of course that's a rather delicate subject. A PTA for the fourth form, well, I don't find it necessary. That's very revolutionary and they don't like to hear it. The dossier test will be removed from next year's programme for the fourth forms, by the way. Because I find that with each subject, although a language is really the best example for this, you keep working towards perfection or at least towards being able to express yourself as well as possible in a number of different areas, with the help of the foreign language. So you shouldn't test that until the very last. That makes the most sense. I can imagine that it might be different for other subjects and for subjects that are concluded in the fourth form, because there of course you get the final examination that year. But with a subject such as English, you shouldn't test that until the very last.*
>
> (2-2-32)

Mark holds a view of language testing which is predominantly summative. Examination marks should be given at the end of a learning process. Drawing up a detailed plan for the fourth and fifth forms takes away the dynamics of the teaching and learning process necessary for teachers to have their learners achieve the best possible results in the sixth forms. Besides, a PTA may have some undesirable legal consequences no one is looking forward to.

> *M: I would really like to dismiss the PTA for the fourth and fifth forms, exactly because that way I would still have control over my own programme, but anyway, that's not possible, because it has to nicely conform to the dossier tests, and so forth. But the more you trust to paper, the more it's etched in stone. You're stuck with it. And learners can, well, ask you for clarification if you don't stick to the programme. And with the current legalisation of society, which is something horrible by the way, everybody knows what he or she//*
>
> *I: I hate that word. Hate that word.*
>
> *M: Yes, can also demand their rights and people are afraid of that, yes, so you feel rather forced. And yes, that's another reason why we're becoming instructors who have to conform to all kinds of agreements, rules, regulations. And well, that personal touch, that rather romantic notion, that you guide a learner from year one to the end, and then in the end you become the assessor, but yes, that notion has come under attack. It just can't be upheld in these days of systemisation and having everything etched in stone.*
>
> (2-2-32)

School facilities

Mark's grammar school has been growing rapidly for the last couple of years. One of the consequences is that his fourth formers are only allowed computer use for fifteen minutes in a row. In the meantime other learners are desperately waiting for some computer access. Besides, Mark feels it is important to have some rooms

where learners can work on their own in silence. He thinks the places the school offers are too limited.

The final restrictive factor Mark mentions is related another of his core beliefs.

Less attention to literary texts

Mark regrets the loss of the central focus on literature in the fourth forms.

> *Yes, you really notice that there is less attention to literary texts. I mean, that's one of the biggest changes for us and definitely for me. We've always felt that literature should be the central point of focus in the fifth and sixth forms. And the skills were taught through that medium. And now, the skills have taken up that central position and literature has been forced into the corner and that just means a complete transition in our didactic approach, for everything. It used to be that we practised reading comprehension through literature, by discussing a story or poetry with them, or at least that was my impression.*
> (2-2-42)

By skills Mark means the attention paid to the four language skills in upper secondary education. Mark feels literary texts are particularly suitable to teach learners how to think and determine and express their positions and views regarding important issues in life. His concerns are primarily ethical. As an example he mentions his preoccupation with First World War poetry.

> *And so you're here, and this brings me once more to the roles of instructor and educator. Through poetry you can discuss important issues that have to do with the way people develop. People learn to think. They choose their positions in life or they start to choose it. They're all grown-ups in the making. Especially in the fifth form. Although that also goes for the fourth. At that age you still have the opportunity to confront them with themselves really, or at least that's what I think, through art, literature. It could also be done by way of abstract texts, but literary texts can be so much more successful. One example was in the fifth form, where I got a lot of reactions from people who said, well, we discussed the First World War in our history lessons, last year. But now it has really become so personal, because I have read that book and I have read those poems and have heard those songs. And that taught me more than all the texts we read back then in our history classes. So there you have it, that's the added value of literature.*
> (2-2-43)

We have elaborated for a while on Mark's evaluation of the second phase reform by way of five restrictive factors that have allegedly affected his teaching in the fourth forms. The factors may help us interpret his definitions of washback on *communicative language education* and on *learner autonomy*.

9.5.4 Washback on CLE

Mark's response to what washback effects he expects the test to have, must be interpreted in the light of what he expects learners to do with the test once it has been returned. His answer, as we have seen, was straightforward. He simply could not tell.

Formative assessment and evaluation and test washback are not on Mark's mind. It seems that his upper secondary classes are moving from one summative test to another. In that process the learners may or may not be aware of what exactly it is they are learning. Mark does not evaluate learning results or the ways leading to them. He partly blames the quinduum system for that, which pressurises the teacher and learners to move on.

The Unicom Final unit eight Test focuses on writing as a target skill. Mark does not see writing as a communicative skill. One would tend to agree with him, given the way in which the learners' writing skills are assessed in the *Unit 8* test, i.e. by having

his learners translate a Dutch letter into English as accurately as possible. Mark, therefore, mentioned no explicit washback effects on communicative language education.

However, when we discussed possible washback effects on CLE, Mark came up with an interesting finding. In the second phase he and his learners speak less English in class than they were used to before the didactic and curricular innovations. This perception might be referred to as an implicit negative washback effect on CLE.

> *Yes. By the way, I also speak English less and less. Well, that's what I've told you before. It's inherent in the system. Unfortunately, yes. Maybe we'll have to change that next year as well, but still. It's rather double, don't you think? Because then you get to the dossier test [test number III. Interview 4]. I did all of that in Dutch, because it's about literature.*
> (3-2-54)

In the second phase the objectives for literary education have changed and have been made less subject-specific. That is why literature and literary texts are generally discussed in Dutch in the revised curriculum of upper secondary education. Mark also points out that Dutch is also increasingly used as the language of instruction in the course materials he uses.

> *Well, you're also dealing with the course materials, aren't you, which poses the questions in Dutch. Not just with literary texts, but also with the regular ones. So then you're inclined to, well inclined, forced really, to also answer the questions in Dutch. You really need a specific assignment to actually speak English, so there you have it.*
> (2-2-55)

There we have it, indeed. Given the fact that the use of English is essential in communicative language education, the increasing use of Dutch may have negative washback effects on the acquisition of communicative competence. We will return to the role of target language use in our discussion chapter.

9.5.5 Washback on LA

Because Mark had already told us he did not see any explicit washback effects of the unit 8 test , we decided to summarise what he had told us before on learner independence. Mark repeatedly acknowledged our summary.

> *I: OK. Yes. Learner autonomy. There was a high measure of learner autonomy here. [Indeed] There were limited possibilities [Yes] to check the progress of learners. [That's true]. You don't really get useful information about what learners know and are capable of until you get the results of one of the Q-tests. [Yes] Then you gain insight into their knowledge, the skills you tested and maybe also into learners' own insight when it comes to doing well on a reading comprehension test .*
>
> *M: Yes, indeed.*
> (2-2-56)

According to Mark, the test may or may not have affected the degree of autonomy of his learners. What remains is that the learners were largely on their own when they prepared for this test. This may have had implicit effects, which simply do not seem to concern Mark.

The final test Mark had selected for discussion, was again a literary test.

9.6 The Practical Assignment literary test

Just before we had our last interview with Mark, he told us about his experiences with a meeting he had just chaired. Mark was mentor of one of his fourth forms. A mentor pays attention to the well-being and learning progress of the class and/or the individuals (s)he is responsible for. In the hours preceding our last interview, Mark had discussed his class and the individual learners. Mark was still so full of what had happened, that it simply had to get off his chest.

At the start of the interview on the literary test, Mark was asked to repeat what he had just confided to me for two reasons. First, it was because we felt his outcry and indignation was illustrative of the context in which my respondent teachers had had to function in 1999-2000. Second, the problems mentioned by Mark were in part related to the so-called dossier tests (Du *handelingsdelen*). The final test I was to discuss with Mark, was originally meant to be a dossier test.

I asked Mark to repeat what he had just told me at the beginning of the interview. As a chairman of the meeting, Mark was first shocked to find out that the adolescent learners of his class were discussed as 'cases' rather than personalities and human beings in their own right. The meeting had focussed on technicalities and regulations, whereas Mark had expected it to be on his learners' well-being and a final assessment and evaluation of their learning results. The meeting was about nine days before the start of the summer holidays. The final school results could not yet be determined, because it appeared that most of the learners had not finished one or more dossier tests or practical assignments. Quite a few learners had either failed to meet the deadlines or had handed in work that was too sloppy to come up to the assessor's expectations. In order to move on to the fifth form, learners have to pass all of their dossier tests. Many of Mark's learners therefore had to do resits of a number of dossier tests or practical assignments at the end of the week, while the school was actually closing down for the summer holidays. Mark feels this event is typical of the irregularities and problems that occurred in the first year of the second phase.

This is how our last interview started:

> *M: Well, based on this very fresh impression that I have just got from this meeting that I chaired as the mentor of this form. You notice that suddenly you have to talk about your children, young adults, whatever you want to call them, as cases that, based on the marks they have produced, have to be dealt with according to all kinds of regulations. And, in my mind, the final outcome will be determined mainly by factors that have been, well, stated on paper, but haven't actually crystallised yet, because these are years of experimentation. The second phase can only be loosely defined. So now we have resits of the dossier tests on June 26th. I'm not really at liberty to say this, and it really isn't good to put it in writing either, but oh well, so they are to retake these Q8 dossier tests, which can still influence their final marks. And now we're making decisions on the marks that we have for now, even though you might end up with different marks in a week's time, and I know of several such cases in my class. So you might have people who, based on their marks at this point, are going to have redo this year, when in fact they won't have to, if they make a 6- or a 6 or even higher, instead of a 5, when they resit the dossier test on June the 26th. So these are rather major glitches, I might say. Yes, for the, well, I don't know if I can use the word, I wouldn't want to say chaos, but rather the imperfections of what we are dealing with at the moment. And the thing you really notice is that you're dealing with so many different teachers that the group that accompanies and leads this number of children is so big, that there is just no time to talk about the actual persons these children really are. (2-4-16)*
>
> *I: Yes, about development. [And the like] And what moves a learner, because that's a subject that is often broached at these meetings.*

M: Yes, you're mainly discussing the learners that need to be discussed, who are the learners that don't meet the criteria. They have, for instance, two fives and one four and then you need to consider whether the subject is a compulsory profile subject of the many subjects they actually have. See, these are matters that could almost be decided by computer alone. And of course you talk about these persons, but mainly in terms of legal consequences. It's all heading that way.

I: The judicial nature of assessment.

M: Which is inherent to the second phase, isn't it?

I: The nature of assessment has changed. Is that what you have noticed in these meetings?

M: You're also dealing with all of the dossier tests, which have to be completed. All the practical assignments, which have to be done. They all have to be completed, because otherwise you're not allowed to go to the next year. And look, of course all of that's important, but it just becomes this tangle of things that you have to think about. That can be quite difficult. I find it very difficult. And then there are also a lot of teachers who aren't really involved in mentoring, who don't know about, well, how does everything work again? How do you move up in the fourth form? That used to be much easier. Now you're moved up based on, well, your choice of profile, your electives, based on the amount of failing marks in your profile subjects, so Maths A1, Maths B1, and so on and so on. Very difficult, also to explain to the parents. So, a huge gap ,really.

I: Yes, so much is clear.

M: You can weave the second phase into a lovely tale, but in practice it's a different story altogether.

(2-2-17)

Mark's passionate outcry at the end of the first year of the second phase, stresses the need to investigate how the second phase context of innovation affects the three constructs relevant to foreign language education: autonomy, communicative competence and assessment and evaluation. We will return to the matter in our discussion chapter.

Below, there are the two sheets that were handed out to the learners when the literary dossier test was introduced. The original Dutch texts can once more be found in appendix VI.

Sheet 1

DOSSIER TEST ENGLISH FORM 4 //////// LITERARY PAPER

In groups of 4 (3 if necessary) you deal with the novel of your choice. You can do this as you think best as a group, but you are advised to do all of the reading **at home** and spend the available lessons on discussing the chapters you have read and actual work on parts of the assignment below.

ASSIGNMENT
Write a paper on the novel that you have read. In this paper, you include information on the following aspects in (small) chapters:

Summary: summarise the story in a maximum of 500 words.

Title: explain why the title is important; state in what way the story is related to the title.

Setting: describe where the story takes place and make clear in how far the setting affects the narrative.

Theme: what is the theme of the story. Explain why you think so..

Main character: describe the main person and indicate whether (s)he changes in the course of the story; if so, clarify that development.

Atmosphere: give information on the atmosphere of the story.

Quotation: select one particular scene/paragraph/sentence and explain why you feel this quotation is important.

Question: think of a question on the novel and answer this question in detail.

Personal valuation: each and every group member gives his/her personal opinion on the novel and comments on the way he group has cooperated.

APPROACH

Every group chooses a leader, who is reponsible for the group. The deadline (the ultimate date on which the project must be handed in) will be indicated by your teacher. Make sure every group member has a copy of the project you hand in. You will be assessed as a group. All of the group members will get the same mark. Therefore, it is crucial to deliberate and consult one another! Arrange how you are going to tackle the assignment. Make sure of an appropriate planning! Discuss what you are going to read/deal with for each lesson. What will you do at home, what at school? The leader ensures that every group member meets his/her commitments. In an exercise-book the leader reports on:

what has been done in a lesson;
what has been arranged for the lesson(s) to come.

It is also possible to ask your teacher questions, which can be discussed with him/her in the lesson to come. At the end of each lesson, the exercise-book is handed in by the leader, so that your teacher can monitor your progress and knows of any questions you may have.

The paper needs to be written in English.

Sheet 2

ASSSESSMENT SHEET **LITERARY PAPER FORM 4**

Group members: 1) (leader)
2)
3)
4)

Title of the novel:

APPEARANCE (10)
(care, division, illustrations, etc.)

PRESENTATION (10)
(language use, style)

COOPERATION (10)
(per student)

SUMMARY (10)

TITLE (5)

SETTING (10)

THEME (5)

MAIN CHARACTER (10)

ATMOSPHERE (10)

QUOTATION (10)

QUESTION (10)

(in brackets, the maximum score for each aspect)

COMMENT: **FINAL SCORE**:

9.6.1 Justification

The final test Mark had chosen was in the end registered as a practical assignment instead of as a dossier test, because he wished to mark the test on a scale of 1-10. Dossier tests could officially only be passed or failed. Practical assignments and dossier tests were new to secondary education in the Netherlands in 1999-2000, even though some schools had introduced them some years before.

Mark sees practical assignments as excellent opportunities to have some more literature in his upper secondary curriculum. Mark's justification for choosing the practical assignment as the third test indicative of his beliefs and testing practice is twofold. First he feels it is an interesting test because it concerns a practical

assignment. Second, the assignment concentrated on what he is interested in most, i.e. three or four learners analysing a novel of their choice on their own.

9.6.2 The knowledge, skils, and insides measured by the test

Mark has difficulty in determining the knowledge and skills that are assessed in the literary project. He mentions 'insight into literary texts' (2-4-23). He feels this insight is dependent on the accessibility of the novel that is being analysed. Mark thinks that insight depends on the degree in which learners are able to relate to or understand what has been read to their own experiences in life.

The groups were allowed to choose from five novels, of which Mark only mentions three. The novels are affordable reprints from literary classics published under the name of *Blackbirds* for educational purposes. The three novels Mark mentions are Graham Greene's *The Quiet American*, Hannah Green's *I Never Promised You a Rose Garden*, and Brian Moore's *Lies of silence.* Mark feels that *Lies of Silence* is the most accessible novel of the three. *I Never Promised You a Rose Garden* appeared to a favourite with many female learners. Mark thinks that Graham Greene's novel was the one least accessible to his learners, mainly because of its Vietnam setting in the 1950s.

> *This test measures//. They work in groups of four, so it's a bit difficult to say this about each individual learner, but I think it measures a number of things, namely insight into literary texts. I put it rather carefully, because a lot depends on the book you read. How accessible are the themes? How far have you come so far, as a person. Let me put it like that. And then again the choice is also very limited, because there are only five books, of which "The Quiet American" isn't exactly a book for kids who've just started to read, I thought. Some of the groups did tackle it. With the help of secondary reading or whatever source they usually do fairly well. It's difficult to see, though, in how far they managed to get the meaning of the novel by simply using what has been premeditated by others, and in how far they succeeded in grasping the information from secondary sources and mingling these with their own impressions.*
> (2-4-23)

Thus, the insight Mark attempts to asses relates to three factors, i.e. the accessibility of the novel that is being analysed, the learner's own personal and/or moral development, and the ability to use and interpret secondary sources to arrive at a better understanding of the novel under discussion. Mark wishes his learners to use sources intelligently and not to copy texts straight from any of the sources they have used.

Mark says he is able to closely monitor activities and progress, because he has the groups write personal journals, in which they report and reflect on their group progress each and every time they meet. It enables Mark to pass comments while the students are working towards their final product. It is important for Mark to formatively assess and evaluate learner progress. All of the group members are made responsible for the final score their final analysis is going to get.

> *The group leader would submit the personal journal to me at the end of each class. In it, I could comment on the things they had written, and it really was a tool, however limited, that gave me the opportunity to see how things were going in the different groups. But you're dependent on what's in the journal, and that's what I have labelled cooperation on the assessment sheets. Even if only ten points are awarded for that. But still, just to show the learners that, of course, this is very important as well.*
> (2-2-29)

Mark also says that the test measures knowledge of what he refers to as the "tools" of literary analysis, such as theme, setting or atmosphere. The learners should also be able to produce adequate summaries. Thus, the literary project links up well

with the way in which the learners had been prepared and tested on the short story *A Day's Wait* earlier in the school year.

Mark is generally pleased with the results, even though he felt that he had been very much on the sideline in the course of the project. He had only been able to monitor his learners' progress by way of the personal logs the groups had produced. Here are some of Mark's impressions.

> *I think there will always be a couple of children who, as it were, become the victims of the other members of the group, and who would have performed better in a more motivated group of learners. On the other hand, you also have these camp followers, who , based on their results, benefit from being in a group of learners fully equipped with what I mentioned earlier. Overall, I'm pleased. There were some exceptions to the rule. Children who submitted real rubbish, but all in all. They also quite enjoyed doing it, in general. I don't know in how far they gave me the response I wanted to hear [Oh, no, but of course] and that will always happen but// Yes, in some cases I was really surprised that they were able to achieve these results. Also because of the way it all worked and you weren't able to get a clear view of the way they worked, because they were allowed to leave the classroom and go to the media room, or work somewhere else. They often worked during other lessons than the eighth period, their English period. And I was just one of the sources they could consult and the one to whom they had to submit their journals. But at times you just feel like you're standing on the sidelines. Especially when you're used to, well, what I often mentioned, being the coach in centre field. But well, that's just a part of it.*
> (2-2-28)

Mark, on the whole, is positively surprised by what the learners managed to achieve on their own. I also asked Mark what his learners should be able to do in English in order to successfully take the test. Mark mentions a 'fair command of English' and at times a 'more than fair command English'.

> *Depending on the difficulty of the book, they have to have a fair command of English. Sometimes a more than fair command of English. And, to be more specific, in the case of I Never Promised You a Rosegarden, they really have to be able to take that step, to plan the, well, let's call it the abstract, into the story, into the two worlds. And that wasn't always easy. But I do remember one girl, who's rather weak at English, but who was in a good group, and who wrote that the other group members were of great help to her, because she could ask them things such as 'What's that about' and 'What does she mean there'.*
> (2-2-30)

9.6.3 Construction and use of the *Practical Assignment* test

Construction

Mark claims the literary test is the result of years of teaching experience. He had relabelled a test format he had been using for years and years, i.e. one in which a learner was asked to analyse a novel on his/her own. A colleague of Mark's suggested adding the summary as a test item. Initially, the ways in which the learners were prepared to the tests was much more guided, directive and teacher-controlled.

> *I: Then construction and use. The third question category. How did you construct this test?*
>
> *M: It's the result of experience. I've done this for years, as mentioned before, in the fourth form. And now we've just relabelled it. I made the subdivision together with a colleague of mine. I had never included the summary, but she thought it was important. Also for the learners, so that they would have the story in a nutshell in their files. .*

I: Was there a maximum or minimum word count?

M: Yes. Yes, but I'm never very strict on that. [Yes, alright] Yes, it really kind of developed over the course of the year. I used to, say ten years ago or so, maybe even longer, treat a book in class, where I was the leading man and all these different aspects of the book, and we went pretty deeply, were treated. Then they'd get a test on it and yes, you were really involved in the process and you would agree with the class to, well, read so many chapters for the next class and to write a little summary on those and then we would start the class with, well, why don't you read your summary and you read your summary and then we would continue. What do we know about the main character, what do we know about the setting, what can we add to that? We basically did the same thing, but much more controlled, much more directed.

(2-2-35)

Mark's learners were given more freedom and responsibility in this assignment than they had been used to. In the section on washback of learner autonomy we will consider in more detail how Mark values this change.

Expertise

Mark says that the knowledge and skills needed to construct this test are the same as the knowledge and skills needed to construct the test on *A Day's Wait*: expertise is all put down to years of practical experience. The test items of the present test do indeed echo the test items of the test we had discussed before.

Conditions

A deadline was set in the introductory lesson of the project. The learners worked on the project for six to eight weekly 50-minute lessons. They were free to use either the classroom or any other room in the school. The teacher was always around for feedback if the learners felt the need to receive any. At the end of a group session, the group leader handed in a reflective group log to Mark, allowing him to monitor progress.

Preparation

The first stage of preparation for the present practical assignment had been the *A Day's Wait* test. The learners knew that literature and the use of literary terms were to reappear in tests to come. Mark regrets the relatively long time span between the short story test and the practical assignment. He blames external and practical reasons for that. Correction of four classes by two teachers was only feasible at the end of the school year, because it would generally take Mark about 45 minutes to score a test of a group of four. In the first lesson, information was given on the two a-4 sheets presented before as Mark's test III. One sheet contained concise but detailed instructions on how to go about test preparation. The other sheet contained information on how the test was going to be assessed. It listed the scores that could be earned for each of the eleven categories: *Appearance*: care, division, illustration, lay-out (10) ; *Presentation*: language use, style (10); *Cooperation* (10); *Summary* (10); *Title* (5); *Setting* (10); *Theme* (5); *Main character* (10); *Atmosphere* (10); *Quotation* (10) and *Question* (10). In the very first lesson the learners formed groups of four, chose a novel and started reading the novel. In the lesson that followed they planned their reading, discussions and writing on their own, with the obligation to report on the activities and progress of each session in a exercise-book called a group log. As compared to the *A Day's Wait* test, no scores were awarded for the individual learner's personal valuation of the novel.

Assessment

Mark started providing details on how he had assessed the practical assignment, discussing the separate items one after the other.

Appearance: care, division, illustrations (10)

The scores that Mark had assigned varied from 2 to10. Mark was also keen on noticing whether the analyses had a table of contents and a list of secondary reading. Outward appearance and illustrations, e.g. taken from the Internet, were also valued by Mark.

Presentation: language use, style (10)

Presentation (language use and style) is not one of Mark's most important aims. The criterion was added to encourage the learners to use words of their own correctly, appropriately and elegantly. Mark's bottom score is a 5 here.

> *M: Well, I think I remember having told them that I did not appreciate reproduced English, that is sentences taken literally from a secondary source. So if they wouldn't give me their own feel to the work, or whatever you wish to call it.*
>
> *I: And for that you assigned varying scores, as was the case with appearance?*
>
> *M: Yes, and with some of the scores I would write that the level of English rather fluctuated. Depending on who was in the group of course. But in any case.*
>
> *I: Could you say more about this, about the language and style? Where do you draw the line and say, well, I assign six points for this?*
>
> *M: Yes, you do that rather subjectively and quickly. But what matters is, that when people succeed in presenting their thoughts on paper, well, in fairly good English, with not too many mistakes, well then, yes, I'm easily pleased. I give them a seven or an eight. If a lot of information is lacking, or they've made some very bad mistakes in their English, I held that against them. But you don't often assign a score below five [No, OK], because their command of English is quite acceptable.*
>
> *I: And in this case language is clearly a tool, isn't it, [Yes, yes, yes], to attain another goal //*
>
> *M: Yes, it definitely isn't a grammar test, you know.*
> (2-4-46)

Cooperation (10)

The criterion was added because Mark's colleague wished to include it. It was there because Mark and his colleague wished to stress the value of peer cooperation and wanted to assess how the learners cooperated in the group. Mark admits this was sometimes difficult to do. Some of the groups mainly worked on the project outside the classroom. In such cases he had to rely on the group logs and on the questions some of the group members had asked him in the course of the project.

If too many arrangements and deadlines were not met by a group, the learners could get a score of only 2 or 3 (2-4-47).

Summary (10)

How to write summaries had already been trained when the learners prepared for the *A Day's Wait* test. Again, Mark felt the summaries ought to include parts of the novel he considered essential.

Mark also mentions two additional reasons why he is partial to having his learners produce summaries of literary texts in words of their own.

> *Yes, the criterion for the learners is really that writing the summary provides them with a better understanding of the story and then in two years, in case questions are asked about the novel, which might be the case, they will feel more confident talking about it.*
> (2-2-48)

The scores for the summaries varied from 6 to 9.

Title (5)

A maximum score of 5 for this item, with a minimum of 1 if the learners had only mentioned the title without any explanation about the relevance or importance of the title. (2-4-49)

Setting (10), *Theme* (5); *Main character* (10); *Atmosphere* (10); *Quotation* (10) and *Question* (10)

So far, Mark's answers had not led to very specific information on the way in which he had scored responses to a particular novel. Besides, Mark seemed to become a little impatient and tended to lose interest in questions for specification and illustration of the ways in which he had scored the separate items. That is why we asked him for any comments on the remaining aspects of the assessment sheet. Mark feels that the items labelled as *quotation* and *question* ask for more originality than the other test items, which he considers as more general and less challenging.

> *I: I think then that I can generalise my other questions [Yes], from setting up and until quotation// question.*
>
> *M: The ones about the setting and the atmosphere are very general. But the ones about quotations and questions are more about originality.*
>
> *I: Yes, those last two. The one about quotations gets ten points and the//*
>
> *M: In the case of "Lies of Silence", someone, or rather a group, would always choose the last sentence, "This time there would be no witnesses". Of course that's an important sentence, but not a very original one. So I would take that into consideration as well. And in the case of the question, well, that also depended on the kind of question, and of course what the answer was. And so there I would assign ten points, but also far less.*
> (2-4-50)

Personal valuation was not scored in this test, whereas the item had scored 10 points in the *A Day's Wait* test. Mark claims personal valuation was assessed indirectly in the scores given for cooperation.

> *Yes, I would usually find out more about their cooperation in that item and of course about how they had enjoyed the project. Because they are the core questions, the ones I often use to discuss a book during a resit or during a literature exam.*
>
> *I: So indirectly, those personal valuations, where you could really find out what someone thought of the whole experience and what they learned from it, they're really also included under the heading of cooperation.*
>
> *M: Yes. Yes. Some kids would only mention the book, but, and then they would mention the cooperation in their afterword or conclusion. But others would, well, include all of that in their personal valuation.* (2-4-50)

Evaluation

Evaluation of the test with or by the learners did not take place. The test was returned amidst an avalanche of marks, tests, dossier tests and practical assignments. Mark told us that, again, there was no time to seriously evaluate the results of the test. Neither did the learners seem to be interested at this particular stage. Perhaps with the exception of a group of four. They seemed a little worried about their modest score of 1.9 and its potential consequences.

> *M: They got everything back in just half an hour, amidst an avalanche of marks, tests, dossier tests and practical assignments.*
>
> *I: Their reaction is often, oh, that's great, or, hmm, that's a disappointment.*
>
> *M: I don't even know, because afterwards we have to get back to our own classroom, because the system rules, to be ready for children from any of the classes you teach, who have questions about the tests or the assignments, there was no one to see me. Actually, one person did, one from the group that got the 1.9. But that person wanted to know what to do with this mark. Whether they would get a mark for the entire year or//*
>
> *I: So they were considering the consequences.*
>
> *M: Yes, a lot of the kids calculate. They don't really look at the actual contents of the assignment, which is understandable at the end of the year.*
> (2-4-52)

9.6.4 Washback on CLE

Mark sees the input and reading of literary texts as incentives to produce writing that has been structured by way of instructions and guided questions. This may lead to writing which is highly personal and in which the learners communicate their thoughts and feelings in a language which is not their own. The learners' written response to literary texts and characters helps them to develop craftsmanship.

> *I: In how far does this assignment, the dossier test, lead them to autonomously or independently learn to communicate in English?*
>
> *M: People who enjoy reading might have learned, a little, how to look at a literary product. Because that's really what you're trying to do here. Not just eating the cake, but also learning what goes into it and how to make it. The recipe and…*
>
> *I: Perhaps they start to get more from a book than before?*
>
> *M: Yes, that's what you hope to achieve, isn't it. The art, the craft even, of writing. But I think that most people, and that's something to be proud of really, are able to lose themselves in a story like this. Well, I find that quite impressive. There are children, or people, who, well, who hate reading. That's a fact of life. Unfortunately, the stimuli for them to read have become rather limited, because they don't have so far to go, because there are only three books left. And they know they can do these in group projects.*
> (2-4-54)

9.6.5 Washback on LA

According to Mark the learners are given more autonomy and responsibility than they are usually given by him in preparation to this literary test. Nevertheless, there

has been teacher control over the activities and tasks the learners carry out in their group sessions. By working with group logs, Mark has been able to monitor progress and signal any difficulties. Besides, the learners can always come and see him if they meet with difficulties they cannot solve or matters they would like to learn more about. The results do not disappoint him. Some of the groups have even surprised him and came up with a quality analysis of the novel of their choice.

Yet, Mark is unsure about any positive washback effects of the dossier test on the autonomy of his learners. What he can at best do is hoping that his learners have learned to appreciate literature in the sense that the works of art are of added value to their own perceptions of life. This is illustrated in the segment in which Mark responds to what learners will do with the test once it has been returned.

> *M: Well, they have to hold on to it, don't they? That for starters. I don't know. I don't really think they'll, well, maybe later in the oral exam . Then they might look at it again, but other than that, I hope they will have learned more about the troubles in Northern Ireland. Yes, some of them mentioned that// actually, wrote: 'Aren't I lucky to have grown up here, instead of in Ulster'. And, for I Never Promised You a Rosegarden : ' I actually did learn something about a different world which is so unlike my own'. So, yes, that's really why I'm doing what I'm doing.*
>
> (2-4-53)

Yet, once more, if there are any washback effects, they remain hidden for Mark. Thus creating positive washback on learner autonomy is not an explicit goal of his teaching.

In view of developing learner autonomy, Mark informed us about an important dilemma. The dilemma concerns the extent of teacher control and his role as a teacher in promoting learner autonomy. On the one hand Mark acknowledges advantages of having learners work on their own, and refers to the sometimes more than reasonable results. On the other, Mark realizes that expert teacher knowledge is essential and ought to be transferred to the learner in one way or the other.

> *I: You see some positive effects in the sense of, as compared to you really directing your learners and have them come up with more than satisfactory end results, it's also of value to have them select and use their own sources [Yes (emphatically)], of which you yourself are one.*
>
> *M: Yes, yes. Although I have less insight into how they're getting somewhere. And that, of course, is the problem I've been wrestling with, so to speak. I spoke about this matter with a professor friend of mine. In how far can you just leave it all up to the learners, so to speak. As in, well, an author is being discussed and the students have to find a relevant scene from the novels or the poems and give their interpretations. That will take quite some time. That's not the right word, my friend said. He once told me he had attended one of these sessions where students were working on their own for about one and a half hours, and afterwards he went back home and wondered, hey, this wasn't discussed and that didn't come up, and I know so much more than what was actually discussed in the session. And of course it is a matter of balancing pros and cons. I once witnessed a class of his about Emily Dickinson. I found what happened there fairly telling. This friend of mine really knows how to tell a story, but this time two students started with the introduction, and had prepared this at home. I was sitting next to another student and what the two students were saying was pretty good and relevant as well and the student next to me was taking notes. After about half an hour, the professor took over and started his interpretations. And, believe it or not, guess what the student next to me did? He crossed out the notes he'd taken from what his fellow students had said and then started to write down what the professor was saying. And I just thought that was, well, of course that*

student hadn't experienced the second phase and wasn't yet that independent. But it's really complicated, isn't it? What is most important?
(2-4-38)

We check for Mark's interpretation of the incident he had just reported.

I: But why do you think the student did that?

M: I don't know, I don't know that student. I just happened to be sitting next to him. Maybe it was also because he was a really obedient student, a follower, someone who reproduces things and doesn't actually think for himself.

I: As in, I'm sitting an exam soon, and then they expect me to//

M: And then I have to, yes, exactly. But of course, that's always the issue, isn't it? In how far you leave out things you feel to be relevant. Because time has to be spent on other matters. But literature is of all times. You can never stop talking about art, about literature.
(2-4-39)

Besides Mark's appreciation of literary works of art, he indirectly refers to a dilemma that educators have to face: what can learners discover and develop on their own and at what particular moment do educators decide to add their own knowledge, skills and understandings, so that mediocrity is more likely to change into excellence. A challenging issue, which touches on concepts such as the zone of proximal development. Yet, we will have to postpone any discussions for the time being.

We will next discuss how Mark's core beliefs, which we presented in chapter 7 and summarised at the beginning of this chapter, are reflected in his assessment and evaluation practice.

9.7 Mark's core beliefs and his assessment and evaluation practice

In the previous chapter on Joy's tests we stated that our first data analyses soon showed how important the role is of a teacher's core beliefs in relation to his/her evaluation practice. Evidence of Mark's professional views appeared to converge in his language tests as well. In this one but last section, we will report on this evidence by mentioning how Mark's convictions are reflected in his testing practice. We will again do so by first referring to important general characteristics of Mark's assessment and evaluation practice. After that, we will more specifically discuss each of his core beliefs in relation to the analyses of the tests and the ensuing interviews on them.

9.7.1 General characteristics

All of the three tests selected by Mark provide convergent evidence of his core beliefs in the sense that:

- two of the tests are related to knowledge of literary notions and the ability to analyse short stories (test 1) or a novel (test 3) with increasing autonomy;
- test 2 was an example of the regular paper-and-pencil tests that were administered every five weeks, and which focused on grammar, lexis and reading comprehension. The test is illustrative of how Mark sees the grammatical and idiomatic foundation needed to be able to discuss literature orally or in writing. It also shows how Mark prefers to

test grammar and lexis, which is fundamentally different from Joy's and Pete's approaches. Mark meticulously follows the selection and gradation of grammar and lexis as present in the course materials he uses. Grammar-translation is one of the prominent features of these tests, in addition to the regular reading comprehension tests.
- Mark believes in initial teacher direction and control.

9.7.2 Core beliefs in the tests and interview data

We see that the three tests have provided convergent evidence of fourteen of the twenty core beliefs Mark expressed in the first interview. Only partial evidence was found in the case of four beliefs. Finally, two of Mark's beliefs were not detected in his assessment and evaluation practice as it was monitored in this investigation.

We will start with an overview of the thirteen core beliefs of Mark's of which convincing evidence was found.

1. the following main educational goal: 'to have learners leave this school with a fair amount of schooling in literature, next to having acquired quite a fair amount of knowledge and quite a fair knowledge of the English language as a means of communication, both oral and written.' (1-2-3)

Two of the tests Mark selected were literary tests. The knowledge component is prominent in Mark's teaching and testing practice. His fourth-form grammar-school learners were 'schooled' in literary notions and were gradually shown to apply these to short stories, and by extension to a whole novel at the end of the school year. The knowledge component is also prominent in the way in which he deals with the course materials and the corresponding tests. The learners are expected to learn and understand the grammar rules and to master the vocabulary offered in the textbook units by way of rote learning. This grammar and vocabulary consistently return in the tests the learners take every five weeks. Mark stresses a number of times that he wishes his learners to be 'creative' with the knowledge they have gathered and to use it in examples and by extension in actual communication.

2. the value of teaching about literature and analysing literary texts. To this end, Mark wishes to empower his learners to be able to analyse and discuss poems, short stories and novels. He is convinced that by reading literature, one learns about life;

Mark's preference for literature and literary texts has already been discussed above. By teaching and practising the discourse required to analyse and discuss literary texts and by selecting appealing and accessible short stories and novels, Mark has his learners communicate their thoughts and feelings about the themes, motives and characters of the literary texts they have studied. English is used when literature is dealt with, which goes for his teaching as well as his literary tests as far as we have been able to detect.

3. the benefits of ex-cathedra, direct teaching, where an educator inspires his learners in convincing performances;

We observed two lessons of Mark's, which were both characterised by his strong and influential presence in the classroom. The first lesson was a preparatory lesson to a short story, in which his fourth-formers were at first quiet but seemingly uninterested. Mark noticed this after some five minutes and started to challenge them by referring to an area of the town in which he teaches that can certainly be called underprivileged. The area is known for its high rate of unemployment, racial problems

and regular fights and burglaries. Mark wanted to engage his privileged learners to sense and imagine what it must be like to be underprivileged and to live in a dismal area of town. The learners literally straightened their backs and became attentive and responsive in English. Now they were ready to read and fathom the short story that was to be discussed.
The second lesson was completely different, but also illustrative of Mark's presence in the classroom. It was a lesson in which the learners worked with the Unicom Finals course materials. Mark first checked in Dutch whether all of the learners knew what they were expected to do in the fifty minutes to come. The learners had to work individually and in complete silence at the unit exercises, which were predominantly related to grammar, vocabulary and reading comprehension. It immediately became clear that when Mark asks for complete silence he does mean complete silence. For the rest of the lesson, the learners looked at their planners and actually worked on their own in a silence that was almost eerie. They were free to come up to Mark if they had any questions, but no one did in the course of that lesson. Surely, Mark's strictness helped the learners to get some work done in that particular lesson.

4. the teacher setting examples of behaviour, linguistic or otherwise, for the learners to copy;

In the course of the year of data collection, Mark often referred to the allegedly important role of the teacher in initiating learning by setting examples. He wishes part of his surplus knowledge and skills to be transferred to the learners. Mark feels comfortable with his approach of direct teaching.

Mark is also very much in control in his class management, as the example of the second lesson of Mark's we observed had shown in no uncertain terms. Whenever Mark asks his learners to work in complete silence, it is complete silence he is after.

5. transferring the academic approach of dealing with literature he had experienced himself to the English lessons at his grammar school;

As a student of English, Mark had also been taught the discourse required to analyse and discuss literary texts. His file of short stories largely consisted of short stories that were once dealt with in the first years of university English. The professor or reader responsible for the lecture typically initiated discussion and debate. Mark had expressed his preference for such an academic approach on several occasions.

6. explicit grammar and idiom teaching at a highest possible level, regularly tested in translation sentences from Dutch into English;

The unit 8 test illustrates what Mark expects his learners to have mastered in terms of grammar, vocabulary and the translation of Dutch into English and English into Dutch. The translation sentences are generally full of pitfalls that require a good deal of grammatical knowledge and translation skills to do them correctly. Moreover, the sentences may deal with grammatical issues that have been dealt with before, which means that the learners are expected to retain what has been learned before.

7. gradually and systematically building up a learner's literary and linguistic knowledge and skills over the years;

The two literary tests provide convincing evidence of how Mark builds up the learners' knowledge of literary notions and its ensuing practice with a number of short stories. In the last test, the small groups of learners are given more autonomy

concerning the ways in which they will deal with their analyses and discussion of the novel they have opted for.

Again, it is characteristic of Mark that he supervises the learning process and monitors the tasks and activities of the groups by having them write their personal logs.

The learners grammatical and idiomatic knowledge is systematically built up by way of the course materials, where the even units repeat what has been dealt with in the odd units.

These two instances of Mark's structured didactic approach, as well as the convincing evidence of his core beliefs, make it likely that he approaches education in similar ways in the other forms he teaches.

8. the freedom education used to offer him as a teacher, which sharply contrasts with the rigidity of second phase innovations and the egalitarian nature of basic secondary education.

This belief of Mark's directly relates to the belief discussed above. Mark regularly refers to 'the old days', in which there were fewer rules and regulations. This enabled him to realise his core beliefs to the full and go for excellence with his gifted learners.

9. the importance of contact hours in teaching and coaching his learners;

Mark preferably uses these hours to deal with literature or to explain or introduce 'new' grammar points. As we have seen from the second classroom observation, Mark also uses the contact hours to ensure that the learners get some work done and are actually given the opportunity to concentrate. If adolescent learners are given the opportunity to do what they feel like, they tend to be less focused on the tasks on hand and more easily distracted. As we have seen in our discussion of learner autonomy, perseverance and concentration were two important parameters of autonomy.

Mark also feels that he has to transfer his own general knowledge and his knowledge of and skills in English to his learners. He feels it is important that the teacher has the initiative here.

10. ambivalent nature of course materials: on the one hand they are essential in structuring and shaping education, on the other the contents and assignments generally do not do justice to the grammar school learner;

Mark definitely uses the course materials to structure the mainstream of his English teaching, as is shown by the regular Unicom Finals tests he has his learners take. Whenever he feels the tasks and tests are too easy for his grammar school learners, he attempts to make them more challenging. There is some evidence of this in the slight adaptations he made in the Unicom Finals unit 8 test, where he combined useful phrases that had to be learned by heart and inserted some extra grammatical pitfalls.

11. the assessment of the learners' proficiency at the end of the learning process, and not during the process of learning to communicate itself;

Mark feels effective tests are summative tests. Formal assessment and evaluation do not seem to be on his mind. The learners predominantly work through the unit exercises on their own and seem to often check their response by way of keys.

12. the incentives that tests provide for human beings in order to get work done and achieve results (1-2-31).

There is ample evidence of this. Every four or five weeks, the learners have to take a Unicom Finals test and a reading comprehension test. This means that a lot of work has to be done in between these tests. All of the grammatical issues and idioms have to be studied and practiced before the next test, which has already been planned at the beginning of the school year. This is likely to be an implicit incentive for Mark and his learners to get some work done and make sure the units are finished before the tests are actually administered and done with reasonable success.

13. the fact that the scores on the national CITO reading comprehension tests are overrated in the final mark a learner gets on his school certificate;

There is ample evidence that Mark focuses on more than the skills he feels are required by the national CITO reading comprehension exams. Reading comprehension is largely dependent on a learner's intelligence and his/her ability to analyse well. Of course, Mark prefers discussing literature to practising reading comprehension exams.

An implicit washback effect might be that reading comprehension test have been added to each of the Unicom Finals tests. However, the reliability and validity of these added comprehension tests do not seem to match the levels of the professional exams, judging on the fluctuating scores the learners appeared to get on these teacher-selected tests.

So far, we have dealt with Mark's core beliefs that prominently feature in his teaching and testing practice. Yet, there are four beliefs of his that are less convincingly present in his school practice. Partial evidence, which can in some ways be interpreted as discriminating, was found for the following four beliefs:

- a division between junior secondary education (forms 1-3), in which a firm foundation is laid of grammar and vocabulary, and upper secondary education (forms 4-6), in which that groundwork is extended and transferred to a variety of literary texts and tasks;
  This belief refers to the way in which Mark used to organise his teaching and testing practice before the introduction of the second phase. As Mark had predicted at the beginning of the school year, grammar teaching was to continue in the fourth forms. He did not refer to any structural or idiomatic foreknowledge the learners might have acquired in previous years. Yet, the grammatical and idiomatic knowledge acquired was still at the service of studying and discussing literary texts.

- the failure of the recent curricular and didactic innovations of basic secondary education and the second phase. He feels both innovations are responsible for two transitions he very much regrets: a move from teaching idiom and grammar from junior secondary education to upper secondary education and the diminishing role of the teacher, because learners are expected to work on their own most of the time.
  We have already referred to the first transition Mark mentions above. The second
  transition relates to Mark's interpretation that his dominant role as an educator has to be reduced in second phase didactics. The assumption is that learners can do a lot more on their own than teachers generally think they can. This assumption has, rather dramatically as we hope to discuss,

led to a situation in which Mark feels he is no longer allowed to interfere with the learning process in view of the second phase innovations. That is why he feels more or less forced to have his learners work more on their own, losing the type of learners who used to graduate from grammar school with the help of the teacher.

- the shortcomings of tests that come with the course materials, which often have to be adapted by the teacher;
  In the year of data collection, Mark seemed to rely rather heavily on these tests. Two reasons come to mind why this appeared to be the case. First, the didactic and curricular innovations demanded a lot of teachers and learners alike. That is why it was easy to regularly fall back on the course materials that provided structure to a curriculum under pressure. Second, it was the first year that Mark worked with the revised edition of the Unicom Finals course materials. In the first interview he had already claimed that not until you have actually used course materials for one or two years, can you actually evaluate them properly and adapt them if need be.

- the fact that the role of literature has been marginalised in the secondary phase innovations.
  Mark has done all he could possibly do to maintain a focus on literature in his fourth forms. Therefore, one could conclude that literature has not actually been marginalised in his English curriculum, judging by the two literary tests Mark had selected for discussion.
  Even though literature is paid attention to fairly elaborately in the second phase objectives, the fact that learners only need to read three novels in a foreign language and the fact that it is more difficult to deal with the same literary input that foreign language teachers were used to seem to have caused that teacher may feel the role of literature has been marginalised. The marginalisation Mark perceives should be seen in relative terms. It was not mentioned even once by the other two teacher respondents.

Above we have discussed four beliefs of Mark's of which the evidence was less convincing. However, there are two beliefs of which no evidence whatsoever was found. They are the following two:

- the benefits of error correction and having the learners analyse the errors that have been made;
  This was a belief Mark expressed in the first interview. However, in the course of the year of data collection, no evidence whatsoever was found of this belief. Not even once, Mark had referred to him or the learners analysing or correcting errors. It may be related to Mark's excuse for not discussing or evaluating the tests his learners had to take: the pressures of the second phase innovations. The result of it all is that the learners are given full autonomy in whether or not they reflect on a test and analyse and correct any errors or mistakes that have been made.

- the uselessness of the formal tests that he was forced to administer to his junior secondary learners when they closed off basic secondary education. Mark claims he has never administered these.

It is not surprising that no evidence was found on this particular belief of Mark's. First, the focus was on the second phase innovations in the fourth forms. Second, Mark's views appeared to be prophetic. In the years that followed the year of data

collection, basic secondary education was seriously revised and the compulsory tests were abandoned.

9.8 Summary

In this chapter we recapitulated Mark's core beliefs and construct definitions as expressed in the first interview. We then focused on the three tests that Mark had selected for discussion in the course of the school year. The first test we discussed was the *A Day's Wait* literary test. It was a 50-minute test on a short story by Hemingway the learners had not seen or read before. The learners had to apply literary notions such as theme, setting and characterisation to the story. The test prepared the learners for future oral and written discussions of literary texts. This first test that Mark had offered for discussion illustrated well his core beliefs regarding the teaching and learning of literary texts. The second test we discussed was a regular test taken from the course materials, which had been slightly adapted. The *Unicom Finals unit 8* test proved to be an effective illustration of Mark's core beliefs regarding the teaching, learning and assessment of grammar points and vocabulary. By way of rote learning the learners had been expected to prepare for the test, largely on their own with the help of keys. Advanced English-Dutch and Dutch-English translations were part of the test. A reading comprehension test was part of the *Unicom Finals* test as well. The reading comprehension test, however, was not standardised, which may have accounted for the fluctuating marks a lot of the learners got on the reading comprehension tests in the course of the year. The learners took tests similar to the unit-8 test every five weeks. The third and final test we discussed was a literary *Practical Assignment* test. In an effort to maintain his original literary curriculum, Mark originally planned the test as a so-called dossier test (Du: *handelingsdeel*). When it became clear that a dossier test can only be scored with either a pass or fail, Mark decided to rename the test as a practical assignment. Now it was possible to score the test in more detail. The test proved to be an interesting follow-up of the *A Day's Wait* test. Now, the learners were asked to analyse a novel in groups of four. The groups selected the novel of their choice from five novels that had been preselected by Mark. He mentions three of the more popular choices of his learners: Graham Greene's *The Quiet American,* Hannah Green's *I Never Promised You a Rose Garden* and Brian Moore's *Lies of Silence.* The practical assignment was prepared well in the first and second lessons by having the learners select the novel of their choice, form groups of four and introducing them to the requirements and assessment criteria of the assignment. The latter was done with the help of two sheets. After the first two lessons, the learners worked on their own, with the obligation to keep up a detailed group log that allowed Mark to monitor the groups if necessary. It was always possible for the groups to consult Mark if necessary. Mark was pleasantly surprised by the results of the assignment, which the learners seemed to have appreciated as well.

The three tests that Mark had selected were again discussed in terms of *justification*, the *knowledge, skills and/or insights* the tests were supposed to assess, details on their *construction and use*, that is information on how the test was constructed, the knowledge and skills the teacher required to construct the test, the ways in which the learners had been prepared to the test, the test conditions, details on the ways in which the test was scored and graded, and finally the teacher's expectations of what learners (should) do with the test after it has been returned. The final two aspects that were discussed were w*ashback* on *Communicative language Education (CLE)* and *washback on Learner autonomy (LA),* in which Mark explained in how far the preparation, administration and discussion of the test stimulate the learners to learn more independently and responsibly.

In chapter 11, the data from this chapter will be summarised, put into matrices, and compared and contrasted with the data from the chapters on Joy and Pete. Our findings will then be interpreted in view of the contents of the three theoretical chapters of this study in a process generally referred to as analytical generalisation.

Our final data chapter focuses on Pete's language tests.

CHAPTER 10: PETE, THE PROJECT MAN

10.1 Introduction

The last of our data chapters focuses on Pete. We will once more start with an overview of Pete's core beliefs and his definitions of the three central constructs of the study. Then we will discuss the three tests Pete selected in the course of the school year. After a summary of Pete's beliefs and construct definitions that came to the fore in chapter 7, we will discuss how Pete reflected on the three tests he had selected for discussion. The divisions, sections and headings of this final data chapter are similar to the previous data chapters on Joy and Mark.

10.2 Pete's beliefs and construct interpretations as expressed in the first interview

As presented in chapter 7, Pete believes in:

- developing his own materials and/or selecting from authentic materials that are available as viable alternatives to the text- and/or workbooks used in secondary education. He feels the regular course materials do not offer what he wishes to transfer to the learners in his lessons;
- the value of collegial consultation and cooperation from the very start of his career onwards;
- progressive change, which involves critically discussing and challenging educational practices;
- being given a high degree of autonomy by his school management, which gives him the opportunity to put into practice his particular approach to teaching English;
- presenting his learners with thematic projects, in which an array of knowledge and skills have been integrated;
- the fact that his didactic approach prepares his learners well for the national examinations of reading and listening comprehension;
- being critical of the CITO examinations that are used nationwide for reading comprehension, listening comprehension and writing skills;
- being learner-oriented, which requires teachers to engage their learners and partly adapt their teaching to the wishes, knowledge and skills of the learners they teach;
- the fact that the result of tests are important to his learners, even though he is far from convinced of the use of language tests himself and often refers to their limitations;
- carefully and responsibly preparing his learners for the tests they have to take;
- the limitations of giving marks that are "objective" when complex skills such as writing are being tested;
- holistic teacher assessment of complex learner skills without rigorously defined assessment criteria;
- integrative testing, based on meaningful input by way of appealing texts or video materials;
- tests that ask for more than the mere reproduction of e.g. grammar rules or idioms without any meaningful context;
- carefully planning the tasks, activities and test over the school year;

- the limitations of a so-called “native-speaker level” that Dutch teachers of English are expected to aim for;
- engaging the learners when dealing with literary texts in a light and relaxed way. Pete objects to the artificiality of mainstream literature teaching in secondary education;
- the fact that many learners consider reading novels in English as a real struggle;
- the need to bridge the differences in educational approaches between the teachers at his school who teach forms 1 to 3 and Joy and himself;
- approaching the innovations of secondary education in a realistic and positive way.

Pete’s construct definitions

1. An effective written English language test

- is communicative;
- is meaningful;
- has to link up with what has been taught or learned;
- is geared at transfer.

2. Essential knowledge and/or skills in English

- the subject always precedes the main verb in English;
- the ability to express what you want to express in English, without relying on any efforts at translating one’s mother tongue into English, but using what you know and can do in English instead;
- a learner’s insight into whether an English-speaking person would understand the utterances produced;
- a learner has to master certain language functions, such as showing people the way;
- a learner has to able to express opinions of his/her own to an English-speaking person, e.g. on a novel the learner has read. In order to achieve this a learner needs to master some concrete skill, such as:
  - a certain amount of vocabulary,
  - a certain grammatical insight,
  - awareness of the interlocutor a learner is communicating with,
  - awareness of register.

When we discussed essential knowledge, skills and understandings, Pete frequently referred to the vagueness and lack of specificity of what exactly teachers are expected to teach. It does not bother him too much, because the results of his learners on national examinations and tests are generally fine.

Insight or understanding has been defined by Pete as the stage in which you somehow structure your thoughts and everything you know of or can do in English in such a way that you can put your message across.

3. CLE

Pete’s definition of communicative language education echoes his core beliefs and the elements he has already determined as essential knowledge, skills and understandings for the learners to learn.

- the essentials are that one person wishes to put a message across, with another person wishing to understand the message. The message has been put in such a way that it is understood in one way or the other.
- the message must be meaningful to the people involved.

Pete feels that subjects that are raised in secondary education too often lack any relevant meaning. He refers to these as “nothingness” (DU: “niksigheid”) Too often secondary education deals with subjects the learners are unable to relate to, and will never feel the need to communicate.

1. LA

- Pete advocates a gradual approach. Learners in the fourth form are often used to ex-cathedra teaching, and have to gradually adapt to situations in which they are given more independence and choice. Therefore, learning with full learner responsibility (Du: “zelfverantwoordelijk leren”) is a vague term for Pete. Besides he feels that his fourth formers are unable to decide on how to go about their learning, basically because they have not been used to doing so in the first three forms of secondary education.
- The tests the learners take reflect this gradual approach: the first tests the learners are more ‘learnable’ than the later ones, which integratively test complex skills.
- learner autonomy involves learners consulting one another and forms of peer learning. Learners are often more successful in explaining one another about difficulties they experience than Pete is as a teacher.
- leaner autonomy involves the willingness of the learner to learn by doing. Pete feels English is a ‘do-subject’. The example he gives is that a learner has to be willing to master at least 70% of the idioms of a given text. This involves consistently looking up the meaning of unknown words in dictionaries and develop guessing strategies to further expand their lexis and subsequently use the acquired idioms in different contexts.

2. **Washback of tests on learners learning how to communicate in English**

- Pete does not see language tests as tools that influence what learners learn and how they go about their learning.
- The tests that are done at the end of a project only test a limited part of the knowledge and skills the learners have been confronted with in the tasks and activities of that project.
- Nevertheless, Pete expresses that satisfactory marks for tests tend to motivate learners and boost their self-confidence.
- If learners score badly because they have not put any effort into test preparation, he has his learners bear the brunt.
- Pete’s ultimate goal is the final mark they get for English on their secondary school certificate. All preceding tests are seen as preparatory stages to meet that ultimate goal of scoring a mark of 6 or higher.

10.3 Pete’s tests and interviews

As a colleague of Joy’s, Pete preferred to do the interviews on the same tests that she had selected. Therefore, the tests we will report on are again the *Dear Nobody* writing test, the *Little Boy* grammar test, and the *Irish Question* reading comprehension test. Pete was of opinion that Joy’s selection for the tests was as

good as any and he felt it was interesting to focus on the same test for the sake of comparison.

10.4 Test I The *Dear Nobody* writing test

For the sake of convenience, the test is presented once more. The writing test was more elaborately introduced in chapter 8.

Choose and do <u>one</u> of the three assignments presented below:

1. Write a letter to the author of the novel **Dear Nobody**. Include at least the following information:
   a. Tell something about the project as it was 'done' at school.
   b. Tell something about your opinion on the movie.
   c. Tell something about the importance of reflection and discussion of similar themes.
   d. Give examples of other subjects.
   e. Tell something about the situation in the Netherlands.

2. Identify with Helen. After a couple of years you decide to send Chris a letter. Include at least the following information:
   a. Make up a motive for writing the letter. Ask after his present situation (Some 'guesswork'?).
   b. Tell something about how life has been treating you (and of course the child) in the time that has passed. (Feel free to use you imagination a little)
   c. Once more you look back on that difficult period in your lives. Also include the role your respective parents played. Are you still behind the decisions made at the time?
   d. You suggest meeting Chris again. (or suggest the opposite, depending on what you have written down under 'a'. Of course you have your reasons for this.

3. Identify with Chris. After a couple of years you decide to write Chris a letter. Include at least the following information:
   a. Make up a motive for writing the letter. Ask how she and your child are doing.
   b. Tell something about how life has been treating you in the past couple of years.
   c. Once more you look back on that difficult period in your lives. Also include the role your respective parents played. Can you understand better now why Helen acted as she did?
   d. You suggest meeting Helen again. (or suggest the opposite, depending on what you have written down under 'a'). Of course you have your reasons for this.

Before Pete was interviewed on the *Little Boy* writing test, he had commented on item D of the interview guide. It was the question about test washback on communicative language education, i.e. in how far Pete felt the test affected the learners' listening, speaking, reading and/or writing skills in English. Pete told us that the tasks and activities of the project that led up to the writing test contained far more knowledge and skills than the actual test did. As such, he felt the test was not representative of the knowledge and skills the learners were expected to have built up. The choice to focus on writing skills in the summative project test had been merely pragmatic. Pete stressed that he did not believe in washback effects of a test at the end of a project. To put it more strongly, he did not believe in language testing at all. Nevertheless, Pete fully acknowledges that the results of tests are important to his learners. Pete also acknowledged society expects teachers to test and prepare learners for the final examinations (3-2-1).

Apart from the comments discussed above, Pete had no further questions or remarks after he had read the interview guide. We were in for another more than

interesting interview, despite the fact that the interview was regularly interrupted. For four times we had to leave the place where we were sitting, to make room for a class or colleague of Pete's. Notes allowed us to start again where we had left off. Here is what Pete had to say about the *Little Boy* writing test.

10.4.1 Justification

Pete fully agrees with Joy's choice for this test. It would have been his first choice, too. He uses no less than six arguments to justify his choice. The bold print in the segment below is ours.

> *I: Why did you actually choose to discuss the Dear Nobody test?*
>
> *P: Because I would almost like to call it a very attractive project. [Yes] Unfortunately the Scenes series// **(Knock on the door interruption 1, recorded all the way)** Yes, well it's going to be like this the entire time.*
>
> *Yes, UNFORTUNATELY (emphasised), I was saying, that Scenes series has ended by now. At least I can't find it anymore, but it was on the air for YEARS. Just half an hour, at lunch time, on the BBC, [Was it?] and they really showed some good things. Sometimes the items would just last half an hour, at other times they would take up to three weeks and I always faithfully recorded them. And the Dear Nobody item just really appealed to me. I had already done it a few years back, the kids really enjoyed it then and that became obvious during the evaluation of the project, when they said "You really need to hold onto this one" // **(interruption 2: Teacher and class coming in)** Adri, let's leave.*
>
> *Those kids back then said you have to hold on to this one, because it's really fun and well, I also believe that all the four language skills are integrated into a project like that. [yes] It has// these days they also have to do something with the internet and that involves// all kinds of realistic and authentic issues are involved in that. It's about teenagers. Yes, what more can I say? [Yes] I just think that it's a very, very appealing lesson series. [I get your point]]*
>
> (3-2-3/4)

The *Dear Nobody* test is representative of many of Pete's beliefs in teaching and testing, which he had expressed in the first interview. The project on teenage pregnancies was partly based on a BBC programme called *Scenes* and had everything a lesson series should have according to Pete. He offers six reasons why the *Dear Nobody* writing test is worthy of attention. Pete found the project as a whole very appealing. When he had evaluated the project with his learners a couple of years ago, they told him they liked the project very much and asked him to hold on to it. Pete also appreciated that the four language skills were integrated well in the tasks and activities that had to be done. The Internet assignment he also considered a boon. He felt the theme, tasks and activities were both realistic and authentic. A last argument Pete mentioned was that the series targeted on adolescents.

10.4.2 The knowledge, skills, and insights measured by the test

Pete is not very specific when he indicates what exactly is measured by the test. He first expresses his belief that he could easily do without language testing altogether.

> *P: Yes, that final test. Well, you have to test something. I actually often feel that you needn't test anything, but apparently these kids want to get marks and the school wants them as well. So their wish is my command. Of course, it might as well have been a listening test related to this theme, because the project involved a lot of listening. But there's also a lot of writing in this*

project, which is why I decided to focus on writing skills in the final test at the time. Well, as you can see from the assignment, they have to identify with one of the two main characters, either the man or the woman. And if the learners object to the creative aspect – I do mention that a lot, don't I? – to go into that, then I added an assignment that's a little more objective. Where they don't have to crawl into someone else's skin. One where they don't have to show so much of their inner selves. That assignment is more factual, I think.

I: Yes, definitely. So when you made these assignments you thought about your learners?

P: Yes, because it's quite demanding for someone to, well, be creative enough to be able to identify with a man or woman like that. Although the entire project revolves around that, because you're constantly asked, well "What would you do in a situation like this?". In that sense it all leads up to that, but if some learners just don't want to do that, that's fine with me. I don't mind . And so, to a certain extent, the test should also include a somewhat factual and impersonal assignment.

(3-2-5)

Pete decided to assess writing skills. More specifically, Pete expects his learners to use the language 'creatively'. The learners are asked to either identify with the female or with the male protagonist of the story. If they dislike doing that, or find identification difficult in any way, the learner can also decide to answer the more factual test task.

We ask Pete to specify what he means by creativity. He feels that English is like Dutch in this respect. Both languages allow Dutch learners to express themselves in a creative way. In order to do so, learners have to develop some communicative skills. With the help of these yet undefined skills, a learner is able to express feelings and thoughts in English. Pete scores this particular ability as 'content'. In addition he also scores in how far the learner has successfully mastered the 'mechanics' of writing, such as tense, word order and the use of appropriate idioms.

I: So the items that require the learners to identify with someone assess their creativity?

P: They do. Such assignments are comparable to Dutch creative writing assignments. But at this school – and we've been saying that for years – English is really comparable to Dutch. It's just that the small handicap of having to do it in English has been added. So then what do we test? We assess to what extent the kids are able to cope with the foreign language, to what extent they can express their ideas in English. That requires some communicative skills. And consequently, when assessing this test, you don't just look at the contents, which I call the communication, but also at the choice of words, grammar. And all of that then results in a mark. I do divide it into two. I give a mark for the contents and one for the technical skills.

(3-2-6)

We attempt to learn more about what Pete means by 'creativity' and 'contents' and how exactly he assesses these.

I: When would you call someone a good creative writer? What are your assessment criteria?

P: It has to engage me. I am the reader and I do emphasise that. I always tell them that they have to imagine that I'm going to be stuck with about 20 or 30 of these test papers and I have to read them all. If all 30 are exactly the

same, then it's going to be pretty boring after just the second one. So I want them to please try, if they go for assignments two or three, to incorporate something of their own personalities in there. And then I just ask them how they do creative writing in their Dutch class. Come up with examples. I ask them to just make it appealing, you know.

I: You mean coming up with examples that relate to their own lives? OK.

P: So in this case about Chris or Helen, you know, but they can make something up as well. It doesn't have to be…

I: Yes. Any other things that you think, yes, this is an aspect of creative writing? As a reader, and assessor, the things in there just have to grip you ?

P: They have to, indeed. Oh, some learners are very creative and really know how to tell a story, but others are just downright boring.

I: Yes. So that means that in your assessment you differentiate between learners who are original [yes, yes] and those who are boring [yes], and perhaps those who are rather factual and detached [yes, yes].

P: They'll very seldom get a failing mark for the contents, [yes] as long as they remain true to the points I mentioned here (Pete points at the assessment sheet). I also feel that you should take the assignment seriously and when I ask them to identify with Helen or Chris, or let's consider Helen for example, and then they write a letter saying they're Helen's grandmother, then that's just the end of the story. But they know that. And they don't do that. So those four aspects, in this case, they have to be integrated into the assignment. But I think they really appreciate that four aspects are being mentioned in the test item, because then they know that needn't bother thinking about that. (3-2-7)

Pete feels that the learners' creativity should be aimed at engaging and gripping him as a reader and assessor. This will be the case if the writer is an original and a good story-teller. Yet, even if learners lack the originality of gripping story-tellers, they will not easily fail the test as long as they stick to the four aspects that must be included in their response. It is important for Pete that a learner attempts to express his/her real self. The foreign language happens to be the tool for this type of self-expression. The learners' thoughts and feelings are communicated with the reader, who in this case is also the teacher and assessor.

Pete is not too detailed in describing aspects related to creative writing. He suggests creative writing is a skill that can be learned. This is basically done when the learners practise writing their answers to reflective questions such as 'What would you do in situations like these?'

Later on in the interview, we again turn to the question what knowledge or skills are tested. We expected Pete to be more specific now, because we had just discussed how he had scored and graded the test. This is what happened.

I: Now what are the learners expected to know and be able to do in order to do this test successfully? What are you supposed to know? And then, perhaps//

P: That's the difficult part. [Difficult in what way?] No, I don't believe that there's anything explicit// You can't just study something explicit for this test. It's a// We worked on it for months and somewhere along the way we check to see how skilled a learner is in transferring a message to paper. And that - //. The first time you do a thing like that, you can really tell that it's not working. And then when you check again three months later, you can actually see the progress in that area. [yes, yes] But it's a bit vague, I have to admit. (3-2-14)

Pete does not point out any specific discrete knowledge or skills the learners should master in order to be able to communicate. Writing in personal and appealing ways is what learners pick up by doing, which will lead to progress after a couple of months.

Pete once more returns to essential knowledge, skills and insights somewhat later in the interview.

> *I: Knowledge and skills. Yes, implicitly we've talked a lot about what learners have to be able to do.[yes] The things they have to apply to their work and which you will assess, concerning the language technical part that is, right? [yes] Maybe we could get back to what you called creativity, even though you've said a number of things about it.*
>
> *P: Well, over the course of the entire project [yes] you build up a lot of knowledge of the subject. [yes, yes] And all through the project, you're asked for your opinion. [certainly] – [yes]*
>
> *I: Certainly. Yes. So you're challenged to provide your own opinion quite early on?*
>
> *P: Let me see. Well, for instance, after they've seen the first episode, the last question they have to answer in groups is, "Helen makes up her mind to do something. What do you think she is going to do?" [yes] They have to think about that and find a solution that is acceptable to all four members of the group. [yes, yes] And often you'll get, well, I would do this and that. Well, the boys will say, I don't believe any of that. She's going to do something totally different. Unwittingly, I believe that assignments such as these will help them do the final written test.*
>
> *I: You've given your opinion before. The way you think about these events..*
>
> *P: And all of that comes back. And you, and you might recognise [mm] little things, like, oh that's right, we said something like that back then, and back then I though this and that, well, I can use that stuff. [yes] Or maybe, wow, she said that back then and I kind of liked that and now I can use all of that. And of course that could all come back beautifully in that brainstorm session.*
>
> (3-2-30)

Pete feels the most important part is the application of any declarative knowledge the learners acquire in the course of the project to the ways in which they express their informed opinions.

> *I: Yes, exactly, yes. So knowing and being able to do something, concerning their creativity, is there anything that you would like to add to that? --- Giving your own opinion, which is what we talked about.*
>
> *P: Yes, in a way knowledge is involved of course, but for the most part it's applying your general knowledge. [yes, yes] Use those brains and your language [yes] and do something with it. [yes]*
>
> *I: Certain knowledge, by that you mean, well, if you want to read a text on your own, then you're going to have to be able to understand enough words to be able to work with that text in a group setting. [yes]*
>
> *P: If I were to give a talk about a drilling platform then I would definitely have to read up (laughter) on the subject, because otherwise it just wouldn't work. [yes] [OK} – That's the knowledge, but – YES.*
>
> (3-2-33)

In the next section we will concentrate on details concerning the construction and use of the *Dear Nobody* writing test. Because details on the test were already discussed in chapter 8, we will limit ourselves to Pete's specific reflections and interpretations.

10.4.3 Construction and use of the *Dear Nobody* `test

Construction

Pete has been used to constructing tests of his own from the moment he started teaching upper secondary classes and began to develop and teach projects of his own. Years of experience have ensured that it does not take Pete long to produce tests like these.

Later on in the interview, we discussed some of the tasks and activities the learners did in the course of the project. The segment shows that it is not always easy for Pete to be explicit about how and why he constructs particular test questions. Ideas for effective test items tend to come up intuitively.

> *I: The way these questions are constructed, for example: 'Their future is bleak. Explain the bleakness in your own words.' How do you construct questions like this?*
>
> *P: Well, I just come up with them. While I'm reading, I think…. yes, that's a pretty good question.*
>
> *I: Yes. All right.*
>
> *P: Yes, it actually forces them to read that text a bit more intensely than just reading it.*
>
> (3-2-56)

On closer reflection, Pete says that questions like these invite the readers to read the text more closely.

Expertise

Pete again refers to his formative years as a teacher. In the opening interview he had already hinted at the importance of two progressive colleagues of his, who had taken young Pete under their protective wings.

> *I: You say questions like these just come to you naturally, while of course there are lots of people who don't find it that easy at all. Where did you get the knowledge and skills to be able to just come up with such questions? Do you have any idea?*
>
> *P: That's just the way I've always been.*
>
> *I: Yes. From the very beginning you really//*
>
> *P: Yes, but also, phew. Always really, my entire life, things like that. [yes, yes] I never really worry about it. Sometimes, Joy just can't get over it. But, well, apparently it's just something that comes naturally to me.*
>
> *I: Is it insight?*
>
> *P: WELL (emphasised), I do believe that some insight is involved there, yes. But that's been acquired. It's experience.*

I: I see. So that means that in the very beginning of your career as a teacher, you wouldn't have been able to construct tests like this?

P: No, I don't think I would have. No, I don't think so. From the very beginning, I was keen on preparing my own classes. I said so in the very first interview. Yes, that's how I got here and that's how I still work here, and – and - //.

I: But those colleagues that coached you back then, you know, you talked about them quite elaborately, did they test in a similar manner? Yes. So they were familiar with [Always prepared everything themselves] Yes, yes, certainly. And also focused//.

P: And always vague as well! Vague! (laughs)

I: Yes, vague, those are your words, though, not mine. (both laugh)
(3-2-46)

Impressionable, yet with ideas and a personality of his own, Pete started his career at his present school. He further developed his teaching and testing practice along the lines of his progressive colleagues.

Conditions

Pete's approach to assessing the *Dear Nobody* test is unorthodox. The administration of the test was not simply a matter of booking two hours and having the learners finish it within the given time. The administration of the test took two or even three lessons. Pete has forgotten, because he is not really that interested in how long it takes for learners to produce a letter they feel is satisfactory. In the first lessons the learners were given the test sheet, and told 'to play with it a little'(3-2-51). It allowed learners to choose the letter they wished to write and produce a first brainstorm. The test sheets and brainstorms were handed in after the first lesson. This approach enables learners to think about their letters at home and look up any words or constructions they wish to use when they start writing. Pete, however, does not allow any notes taken into the classroom on matters the learners have prepared outside the classroom. In the remaining lesson or lessons, the learners write their letters. Pete has forgotten whether the learners were allowed to use the classroom dictionaries. He feels it is important for the learners to write the letters in a relaxed way, without the pressures often felt by learners when they have to take tests.

I: What were the assessment conditions, which sounds rather technical, but how much time was given to do the test?

P: I think about two lessons. Two lessons, yes. Well, first you hand out the paper and the assignment. [yes] Play with it a little. [yes] Maybe even three lessons, I'm not so sure. But it doesn't matter to me. Then they go home and hand in their papers and then the next lesson//

I: So they're allowed to think about it some more at home? [yes] And you're talking about the test here, aren't you?

P: I am, Yes. And then they come back for the second lesson and then they get their papers back and I tell them that they're not allowed to have any materials from home on their desks, because I don't want to read your mom's or dad's work.

I: No. Obviously, yes. So as far as your own responsibility is concerned, you don't feel like you have to reserve one testing hour for that. [no] I mean, it's a writing assignment [yes] and I'm going to assess it [yes and that's it] and I don't care about the rest. [yes] OK. And the same goes for dictionary use?

The learner decides whether he wants to use one or not? [yes] OK. Well, we mentioned the assessment and scores quite thoroughly. I believe we did, didn't we?

P: Yes, although I do have to honestly admit that the only thing that's objective about it is the fact that I'm the one assessing it. And nothing else. And in this specific case we have a colleague who really doesn't like to work with us. And I do believe that that class was assessed very differently from ours. [yes, yes] Well, I can't prove it, but I believe that if I would have checked those tests, that their marks would have been much lower. But I never have any problems like that with Joy. – [yes]. I do hope that she mentioned that as well, by the way. (laughter). What if she said, that dude, well you just cannot cooperate with him. (laughter).

I: I don't think you'll be surprised to hear that she didn't say any such thing! Yes, assessment conditions.

P: Yes, easy. Relaxed, relaxed.

I: Yes. No pressure, no force. [no, no, no]

P: Don't make it too exciting for them.
(3-2-51)

We wished to learn more about Pete's unorthodox approach to language testing by giving the learners the opportunity to read the test papers beforehand, reflect on them, and take notes if they feel like. Pete approaches tests in this way more often, as long as the procedure does not lead to any rote learning of answers the learners look up in between lessons. Pete is not at all worried about learners looking up difficult words. One of the reasons might be that the learners have been allowed to use dictionaries when they are sitting their final examination reading tests from the year 2000 onwards. The example Pete gives in the segment is below is that he has his learners study texts on which he is going to ask open-ended questions in the actual test. Pete rarely ever uses the same test again in the year that follows. Primarily, because he wants some variation and likes to stay motivated himself.

I: Is it that way with all your tests?

P: Well, we've just happened to have done reading comprehension and I had thought that I would let them study the text an hour before, before they started answering the open-ended questions. In 5 Atheneum we just had a week of testing and there they were allowed to take that text home a month before the actual test. [yes] Like, if you want to look at it, go ahead [yes] and if you don't, well then don't. [yes] Yes, but what kind of questions are you going to ask? Then I say, well you guys know me, so you know I can come up with the most outrageous questions. (both laugh)

I: So just be sure to know what the text says and then you can face any surprise you throw them. [Yes] Yes. And whether you do or don't, it's all up to you.

P: Yes, yes, I don't really care. But you can't do that with multiple choice, because then they can find out where the text came from and then they can memorise all the answers real quick and be done in two minutes. So you just can't do that.

I: Has that ever happened to you?

P: I did have an experience one time where I thought, well, something wasn't going 100 PERCENT THE WAY I THOUGHT IT WOULD. (Laughter) But, oh well, you can't lay the blame anywhere anyway.

I: What do you do in a situation like that?

P: Nothing.

I: Just hand out the high marks and //.

P: There's no other option

I: And the next time//.

P: I'm not bothered much by that either. [No, no] NO, (emphasised) and I won't give them an extra difficult test the next time either, I think that's just so silly. If I make a mistake like that, I've made that mistake, that's all.

I: But you won't make that mistake a second time, so you'll change all that needs to be changed to prevent that from happening again.

P: Yes, I think that, in a case like that, if it were ever to occur, and I do believe it has happened to me once before, it almost has to have.
But then you just wouldn't use those texts anymore. [mm] I hardly ever use tests from the previous year anyway. [mm, no] I have a whole stack of those things, I can select whatever I need [yes] and then sometimes I'll make some changes. [yes, yes] ---

I: Why is that? That you feel like you don't want to//.

P: Because it's the same type of test. [yes] And to watch Dear Nobody year after year, well, I wouldn't fancy that very much.
(3-2-52)

We will now turn to the ways in which Pete has prepared his learners to the test.

Preparation

At the start of the project, the learners were given a five-page document with assignments and texts. In chapter 8 on Joy we had already discussed three examples from this document. Because Pete went into test preparation so elaborately, we present all of the eight assignments in full. The sheet was already in English, and has not been translated.

Dear Nobody- a project for class use. Book written by Berlie Doherty

A. A fragment from: the Snapper
Read the extract carefully. In small groups try to decide on how you would 'act' the different parts (How old, for instance, do you think the twins are?).

B. Read the text on teenage pregnancies carefully, then answer the following questions:
Par. 1: 'their future is bleak'. Explain this bleakness in your own words.
Par. 2: explain the figures mentioned in this paragraph.
Par. 4: explain the statistic argument.
Par. 5: what is the function of this paragraph?
Par. 7: 'what caused it'. What is 'it' and 'what' caused it? 'This pattern'. What pattern? 'Plummeted'. Guess the meaning of this word without a dictionary – use context instead.
Par.9: 'a more likely reason'. For what?

C. Watch the first episode of **Dear nobody** . Then discuss the following points:
1. Was Helen wrong to keep Chris waiting for so long without an explanation?
2. What do you think of Chris's reaction to the news?
3. Why did neither Helen nor Chris tell their parents?
4. What courses of action are open to Helen and Chris? Which do you feel is the right one?
5. At the very end of this episode. Helen makes up her mind to do something. What do you think she is going to do?

D. Discussion about the issues raised in 'C'.

E. An extract from the novel.
Change this extract into a 'script'. What do you leave out, how do you show Helen's feelings, what other difficulties are there when you change a novel into a film?
Try to see through the 'tactics' of the director by evaluating the first episode.

F. Watch the second episode. Then discuss the following points:
1. What do you think of Mrs Garton's reaction when she hears that Helen is pregnant?
2. Discuss the way in which the abortion issue is dealt with in this episode.

G. Watch episode three. Then make your **personal** notes about the following discussion points:
1. Do you think that Helen's family adjusted well to Helen's decision?
2. Does Chris have a responsibility as a father now that Helen has broken off their relationship?
3. Why would Helen suddenly decide to send all her **Dear nobody** letter to Chris?

H. Watch ***The Snapper***, a film based on the book of Roddy Doyle.
Think about different ways of approaching the same theme.

On the next pages of the hand-out, the learners find the texts they need in order to do the tasks and activities. In the paragraphs below we will present a number of segments that illustrate the activities Pete undertakes to prepare his learners for the writing test.

Because the project sheet played such an important part in preparing the learners for the test, we discussed it in more detail. Pete shows that he has clearly thought about 'staging' the activities. He starts with an extract from the movie *The Snapper*. The language input and variety of tasks are impressively balanced. Pete is in favour of selecting extracts that are humorous and present unexpected turns. The theme of a project has to relate to the learners' interests. The language input can be as varied as it has been presented here, with a movie extract, a play, and two informative texts on teenage pregnancies. The language output is varied as well. The learners have to speak, express and discuss their views, write down their answers, and do reading comprehension and listening assignments.

> *I: You were saying that regarding assessment, it's hard to point out specific knowledge and skills, weren't you? [Yes, I was] over the course of your career. In fact, you've always found it easy to make a test. [Indeed] you had a certain knack for it [yes] and you've basically continued along that path. [yes, I have] Yes, right. Well, regarding how the learners were prepared for the test, perhaps this is a good moment to have a look at Dear Nobody, don't you*

think? [yes] The products for class use, [yes] you started with a movie extract from The Snapper. [yes] Great movie! [YES (emphasised)]

P: That's the part where the daughter tells them// [yes] that she's pregnant. And I, I wrote that out, and then they had to perform a little play with that, right? Ah, yes, that was it, right. Well, ehm, I sort of wrote that erm// the first chapter of the book as a kind of screenplay. [yes]

I: Let' s see. – Why did you choose this opening for the project?

P: For laughs! It's totally crazy of course. There's this chick, who says she's pregnant, but she won't say who did it and these, these kids in between, looking back, I should've cut them out, those little kids, because they're a bit disruptive, but they're in the film extract. And, and it's just a lot of fun to play act. They also find it quite odd to do. It basically tests their reading skills. It requires them to identify with these characters. It tests their speaking skills and it's, it's basically just a really good appetiser. It's really quite good.

(3-2-48)

Pete is quite active as a teacher in the project. It is not a matter of leaving the learners to themselves. Observation allows Pete to interfere when he feels events or texts are interpreted without giving them much thought. The fragment from *The Snapper* has to be enacted, with Pete being a kind of stage director. Pete feels drama exercises like this one help learners to identify with the characters. By doing so the road towards the *Dear Nobody* writing test, as well as to the writing tests to come, is carefully paved (3-2-49).

Pete skipped activity E, in which the learners had to change an extract from *Dear Nobody* into a script. He considered that an atheneum assignment. It was too complex for his havo learners. Pete also mentions a boon for his havo learners. The project informs them about two novels they can put on their reading list if they like. He refers to learners reading the novels as 'digesting' the novels' (Du: 'De romans verwerken'). In the following segment, Pete explains how he tackled test preparation, starting with the hilarious and engaging opening activity.

I: Well, what you said is important, isn't it, that an introductory activity should captivate, engage. That's very important for a project. Or would you// [Yes] OK, yes.

P: Yes, if in the first class they would think 'Oh my God, is this what we're going to be doing for the eight lessons to come', I might as well go home (laughter). Well, B, that was a difficult text. It includes some factual things to help them identify and it's also just a bit of reading comprehension. Then it's on to the first episode, after which they do some talking. First they do it on their own and then they do it in groups. Then they get that piece that refers back to A, because they have to, they have to/. I don't think we did this part, by the way, E, I mean. Because, I really think that's an atheneum assignment. [I see] And it might even be fifth form material. So I'm afraid I skipped that. Well, then we watch the second and third episode and then we watch The Snapper.

I: Yes. They actually saw the whole film?

P: I believe they did, yes. And I believe that afterwards they read two books in one go again, and they can just add them to their lists.

I: So they get that as well. Sounds like a good deal. Although reading novels is no longer as important now, in the second phase, as it used to be, isn't it?

P: No, it isn't, but they still have to digest some, so it's quite convenient that they have already digested these.

(3-2-50)

The Dear Nobody test is the second step in preparing the 4 havo learners for their final examination writing test at the end of the year. The first step was the *Little Boy* grammar test, which will be discussed later. The examination writing test will resemble the *Dear Nobody* test. It will also relate to a project, for which Pete and Joy will have only four lessons.

Pete says he always advises his learners to pay attention to the following aspects related to the mechanics of writing:

P: I always tell them, just check your verb forms one more time. Just the verb forms. Have you used your tenses consistently? If you start something in the present tense, then KEEP IT (emphasised) in the present tense. Is there enough variation, in the length of the sentences, in the verb forms?

I: What do you mean by variation in the verb forms?

P: Well, if it's only simple present, it doesn't make for a very nice read.

I: So you do expect your learners to //. On the one hand you say be consistent with the tenses you use, but on the other you would like them to vary their tenses stylistically? [Yes, yes] OK. So we have the components of content, grammar and idiom, you mentioned a bit of spelling. [Yes] Choice of words and the construction of the sentences. [I did,yes] And as an example of construction you mentioned how you advise your weaker learners: 'Start with the basics'. [Yes, I would think that would be the easiest thing to do] Be sure to //.

P: Yes, look, in fact that's how you start out in the first form and then you hope that they will have got better in the fifth or sixth forms. A child learning Dutch will also write 'And then we went on the bus and then we got an ice cream' in primary school and in the sixth form of high school, this just isn't accepted any more. And so the same goes for English.

(3-2-10)

Pete regularly advises his learners to have a look at the verb scheme practised and tested at the beginning of the school year, and to vary tenses to suit the style of the letter. Whether or not the learners take up his advice is largely up to them. As long as the learners seriously do the tasks and activities of the project and learn by doing, they are given a lot of freedom by Pete.

Foreign language learners learning how to write need to master vocabulary. Pete is critical of the ways in which vocabulary is generally trained and tested at secondary schools. His alternative is varied and regular language input on themes that appeal to the learners, similar to the authentic input of the texts that were studied in the course of the *Dear Nobody* project.

I: What were your concerns again, regarding the idiom test, where people had to learn six pages of vocabulary, or even more, by heart?

P: Well, you might as well tell them to get out their dictionary and study six pages. I just don't like that at all.

I: Yes. No learning effects?

P: No. That's just so far removed from reality. I find that you should see a language as a tool to achieve something, like communication. And you want

to tell me something. And if I test you on six pages of idiom, you don't have anything to say to me. Then all I'm testing is whether you can remember something, aren't I? That's all I'm testing.

I: Right, short- term memory and literal reproduction.

P: Yes. If you ask them a week later, they won't remember a thing. They often enter the class with their books open, saying oh right, this and that and that, close their books, write everything down as quickly as they can, leave the class and they don't remember what on earth it was about.

I: How do you try to make sure that your learners do in fact learn enough vocabulary? That is, after all, one of the components you wish to assess.

P: I call it input. The more you expose them, the more you have them read, listen, watch, and the more engaged they are, which means the more attractive the material is, the easier that is.
(3-2-12)

Input of authentic texts, scenes, and a film had indeed been a crucial part of the project. We asked Pete whether speaking or writing English in the classroom played any role in preparing for the test. We asked the question, because in the two lessons of Pete we observed, he appeared to speak Dutch most of the time. The learners, however, were expected to do the project tasks and activities in English. Pete says that all of the assignments deal with putting information across in English. In his answers he refers to dictionary use and the strategies of description or paraphrase, which may help the learners learn to communicate in English.

P: Yes, well, all right, all of the project assignments incorporate something that the learners have to try to bring across. [yes] And whenever they don't succeed, I'll have them reach for the dictionary to look up vocabulary they need to use. [Indeed] Although I often tell them that it's actually better to just describe it.
Because often you'll get this really complicated word in a very simple sentence and stylistically that just doesn't look so good. So describe what you want to say OR (emphasised) just don't use that bit. It's your English.

I: So dictionary use is involved as well. Someone who's very capable of using dictionaries [yes], will probably score better and can better decide which words to use, or is that taking things too far?
Is there any training as to how to use a dictionary?

P: No, I find that an automatism. No, it isn't// Not at all really. No, they just use it.

I: It concerns the dictionaries you have here at school?

P: Yes, but they can also bring their own dictionaries, I don't mind.

I: What dictionaries are they?

P: I've got all kinds, English-English, English-Dutch, Dutch-English.

I: They can choose from them and use whatever they want, yes.
(3-2-13)

Pete sees dictionary use as an automatism. The role of dictionary use and vocabulary when learners learn to communicate in English will be further discussed in our final chapter.

The only time the learners' writing skills had been tested before the *Dear Nobody* test was a book report that was scored and graded in the way Pete always scores writing tests. It goes to show that the learners could have got used to the way in which he commonly scores and grades. The segment also illustrates the autonomy he gives his learners in either taking up his advice or discarding it because 'the weather happens to be nice'. This sense of realism will be welcomed by a lot of learners.

> *P: A book report once, and I assess that in exactly the same way. So I'm rather vague in assessing. I also jot down here and there//. I'll write that you translate a lot directly from Dutch, be wary with that. Things like that.*
>
> *I: And consequently, you expect a learner to remember that and keep that in mind when he's working? And//*
>
> *P: A number of them will do so. They, they//.*
>
> *I: Do you compel them to do so in any way?*
>
> *P: No, no, I hardly ever do that. No, I think they should make up their own minds.*
>
> *I: Right. So there might be students who think, well, that's just fine and dandy my dear teacher. [Yes] but?*
>
> *P: Yes, but the weather's too nice, I'm not going to do a thing! Well, I don't mind. YES, HA HA (laughter) [YES, YES (laughter)] And right you are, I think.*
> (3-2-16)

In the course of the project, the learners have acquired or learned details about Dear Nobody and the theme of teenage pregnancies. Besides, they have practised tasks and assignments in which they had to identify with a character or the situation a character was in. The learners got used to giving their private opinions on matters and discuss them with their peers and with Pete himself. This way, the knowledge the learners need to do the writing test successfully was gradually built up in the course of the project.

Assessment

Pete scores and grades towards two separate marks: one for content and the other for lexis and grammar. After a first correction, Pete first puts the tests into piles. He then intuitively gives the two marks for the tests, which he writes down in pencil. After a day or two, he again has a look at the scores he had given and compares a couple of tests from the various piles. He also tends to consider what learners have previously scored on a test, all in an effort to score consistently and reliably.

> *I: That brings me to the subject of how you assessed the test. You give them two marks, one for content and the other for lexis and grammar. How do you go about assessing a test like this? Do you read it first, and while you're reading their work, which has to engage you through its content and creativity, do you score the test in any way, or do you keep these matters separated?*
>
> *P: Yes. Yes, I do. [yes] And after having checked all of the tests, suddenly I find that all throughout the room little piles of paper have appeared, depending on the number of learners.*
>
> *I: So after reading it you think, hmm, to me this means that// I'm putting this one on that pile?*

P: And then I'll give it a mark in pencil and then I just//. Then I just let it rest for a day or two [yes] and then I come back to it and think, gosh, I gave this learner a seven for the content whereas I gave this other one a five, [yes] let me check to see if that's right. [yes, yes, yes]

I: And so is that the first mark you give them, for the content? And only later you look at//.

P: NO (emphasised), I give them those marks, but only later do I check if I did it right. I usually also get some, some - //. It's pretty much impossible now because we, or I, moved up the writing test. Usually for the havo learners, it's the first school exam that they have in the fifth form, and now it's still the first, [mm] but they get it three or four months earlier than before. [yes] But usually I'll grab some from the previous year and go through them again, [yes] and those results basically decide on how I grade the new ones. [yes, yes]

I: But these piles that you place on the floor, do they already have both the marks on them? [yes] That distinction that you make, well for your language skills you get this mark and for the content you get this other mark [yes, yes]. That's interesting, isn't it. The process that you have come up with to check writing tests. Writing can be very//.

P: I, I have to honestly say that I believe that a large part of it is based on intuition. [Mm, yes, of course] And I'm not embarrassed to say so.
(3-2-18)

Pete admits that his assessment is largely based on intuition and experience. Nevertheless, his concern with reliability issues is sincere, as can be illustrated by his efforts to improve inter-rater reliability.

I: Of course it involves a risk. What if someone, a teacher, isn't very skilled in the language or what if he focuses mainly on grammar//.

P: Yes, but that doesn't happen with us, because you know that Joy and I work together. We exchange each other's work. That's one way that we try to achieve objectivity. So I //. That risk is almost negligible with us.

I: Well, I'm not saying it isn't. I meant it more in general.

P: Oh yes, of course that's true.
(3-2-20)

Pete illustrates his concern with reliability and fairness with three other examples. The first example is his claim he has never had a single complaint of learners who wanted to have their work assessed by another teacher in his 25 years of teaching experience (3-2-24). Another example of the reliability, and perhaps even validity, of his test and marks is implicit in the following event. Recently he met a former learner of his, who was now studying English at university. Pete asked him how he had been doing in his class, and how he was now doing at university.

The funny thing is, just last Sunday at the station, I talked to a former learner of mine, who's now studying English at university, and I asked him if he was any good when he was in my class. What kind of marks were you getting in my class? Well, they appeared to have been fair enough, mostly sevens. He wasn't even very good. [I see] He said, well, I'm doing pretty well now, so you just keep on doing what you have

always been doing, because my present marks are MORE (emphasised) than satisfactory. (laughter)] (3-2-27)

A final illustration of Pete's concern with reliability has been taken from the end of our interview on the *Dear Nobody* test. We asked Pete whether he wished to add anything to what he had already expressed. Again Pete illustrated how concerned he is with reliability and fairness. He told me how difficult it is for him to decide or determine what exactly a person knows or is able to do when (s)he has a mark of, say, a 7 for English.

P: No, no, my// Yes, actually what always troubles me with this type of test is, well, I've graded this with a seven, but is it really a seven? I find that difficult. It's VAGUE (emphasised), I think .

I: Would you like to have given that learner another mark?

P: No. No, you wonder if that really is the level of a seven, like say if it were to be checked in Alkmaar instead, or something. [mm]

I: Well, no. You already know the answer.

P: HMMM, (emphasised). I wonder if the answer's really no. It doesn't have to be. But it's, it's, yes I find that tough. You don't want to give them failing marks, but on the other hand you don't want things to get too crazy and it has to stay fair. And really, yes, I do believe it's fair, but still. Well, I just ask myself that question a lot. Questioning my own abilities.

I: Another trait that//.

P: Well, it's a pretty good one really, because thinking that you always know everything isn't good at all either. You have to pretend to know everything towards the kids. Yes, you shouldn't say, well guys, it says it's a seven, but I'm just not sure, then they'll start whining. [yes, then] Yes, then you're in for it. So they'll never hear my insecurity. But I check those tests according to my best ability, and I do use tests from previous years, to get into it, and I put them in the different piles and compare them and that seems pretty reasonable to me and then at times I exchange tests with Joy, and so on. [yes]

I: Which means your scores are comparable/

P: Yes, and they should be. It shouldn't be any different being in Joy's class. In atmosphere, yes, but not in marks. They should be the same. But I think we both strive towards that.
(3-2-60)

We will finish the section on construction and use with some additional details on the ways in which Pete has assessed the writing test.

We already mentioned that Pete marked for content and that he gave a mark for lexis and grammar. The marks were weighted differently. The mark for content has a weight of 1, and the mark for vocab and grammar a weight of 2. Thus a learner scoring a 7 for contents and a 4 for the mechanics of writing gets an insufficient score of 5. Pete has also set a bottom mark for this first writing test. No test taker would get an average mark lower than 5 for the *Dear Nobody* test (3-2-8)

In the course of the interview Pete gives examples of the kinds of grammatical mistakes he objects to. He dislikes errors in finite and non-finite verb forms, such as the distinction between infinitives and past participles, e.g. 'go' versus 'gone', lack of

knowledge of irregular verb forms, the –s that is missing in the present tense third person singular, and certain direct translations from Dutch into English, such as 'I want that he does it' for 'I want him to do it'.

> *P: I have to understand what it says. I must be able to assess how skilled they are. The simple things annoy me, you know, like when they can't use 'go' and 'gone'. Then I think, blast, they should have learned that by now.*
>
> *I: And then you're talking about the infinitive and past participle.*
>
> *P: Yes. And then when they say 'He goed' instead of 'He went' or something like that, you know. It doesn't really happen that often, although they do mention I 'thinkt' sometimes. And that just bothers me.*
>
> *I: Yes, so those are mistakes that you're strict with?*
>
> *P: Yes, that, - that would – that's one of the few things where I draw the line. And then I say, look, it's just too disruptive there, you should have known that. This, you should have been able to use this. I also think that a learner should be able to put an 's' behind the verb when you're talking about 'he'. They've been taught that for three or four years, and I even repeat it, so, yes, they should be able to finally get it right. So the basics, which you ought to have have learnt in the junior secondary forms, you should be able to get those right. And when they want to say "I want him to do it", they often use "I want that he does it". Then I think, well, I do make a note of it, but I'll think well, that's Dutch, and so we still need to get rid of that and so we'll try to do that in the coming one and a half years that we still have together. To straighten that out.*
>
> (3-2-14)

Pete does not pay attention to each and every grammatical mistake. He tends to be more lenient when it concerns grammar that is more complicated. It is difficult to determine, though, exactly what Pete accepts and what he scores as errors or mistakes. Pete also mentions that he objects to word order errors.

> *I: Are there any other things, in the grammatical area, that you consider to belong with the basics?*
>
> *P: Yes, word order.*
>
> *I: Word order, OK, yes.*
>
> *P: But really those are the only two things that are different in Dutch?*
>
> *I: -Yes, yes. You – mean verb forms – [yes] and word order? But what about things such as, just to give an example, adjective and adverb, you don't really pay attention to those?*
>
> *P: NO (emphasised), I don't find them all that important.*
>
> *I: You let them get away with it?*
>
> *P: Well, you read it more and more often in English, that they just leave that out, one way or the other.*
>
> *I: And differences between American English and British English and this type of English you get. So we'll just say verb forms and word order, you watch out for those two. And they determine --//.*

P: They pretty much determine the lexical and grammatical mark.
(3-2-17)

The position of adverbials of time and place is another category of error Pete pays attention to.

I: Well, we already mentioned vocab, didn't we? And well, the construction of sentences, does that make you think of word order, which you already mentioned? Or forming complex sentences?

P: Because those are really the things, [yes] that I warn them about, right. [yes] If you put the adverbials of time or place at the beginning of the sentence, then in Dutch, we just turn the whole thing around. And yes, I do tell them to be careful with that, and I'll say to my 4 havo learners "Just put them at the end of the sentence". Then you're always safe. But when they say "Yesterday went he", for example, that aaaargh, darn it, dummy, don't do that, because I warned you so. So those are really the things that I've mentioned in class, things that I expect them to do correctly.

I: Yes, certainly. But what if someone says, well, that adverbial goes at the back, right, but it also adheres to a certain order, and so if they say 'He went to //.

P: Yes, you mean place before time. Yes, all right, [you're lenient with that] Well, I don't mind it once. I do believe it's in the grammar review, so they should really know that, but I don't find it VERY (emphasised) annoying.

I: Yes. So that means that you differentiate between inversion erros you don't want to see, 'yesterday went he'//.

P: I find that more important than the other thing. Although, if I were to see 'He went in 75 to Amsterdam'. I believe that I would make a note of it.

I: You mean it sounds a bit odd.

P: Yes, I do find it odd.

I: And say it occurs three times, then would it start to annoy you as well?

P: Yes, then it would annoy me. HA, HA, HA (laughter).
(3-2-29)

Finally we turn to Pete's assessment criteria for appropriate use of vocabulary. Pete dislikes disturbing repetitions of words. It is not entirely clear how exactly he assesses these 'errors'.

I: Choice of words, idiom? Could you say more about that? How big of a role it plays in //.

P: Yes, I would almost say that it's like a threshold-level thing. You notice it. A child who only uses words with one syllable is rather, rather [yes] limited. And if they only use the word 'nice' and they can't make it more varied, [yes] they're limited in their choice of words. [yes, yes] In Dutch they call it disturbing repetition. [yes] I often refer to that. [yes] That they should try to avoid that and that they should be able to notice it themselves. [yes]

I: Disturbing repetition or using a particular adjective, instead of the many available alternatives that can be used. [YES (loud)] With 'nice' being a good example. [yes] Yes.

P: I often circle things like that, you know.

I: Meaning that it's not really wrong, but//.

P: In one paragraph you get about ten 'man, man, man, man, man's'. Now couldn't you have come up with something better than that? [mm]
(3-2-25)

So Pete distinguished between content and creativity on the one hand, and the mechanics of writing on the other. In other words, he differentiates between the meaning and style and the linguistic competence of his test takers, without becoming very explicit. This means that his assessment is holistic to a great extent, of which Pete is aware. Nevertheless, he is genuinely concerned with issues of reliability, fairness and validity, but never sees further specification of assessment criteria as a solution to further improve the essential measurement qualities of his test.

Next, we will see how Pete interprets evaluation in relation to the *Dear Nobody* writing test.

Evaluation

Pete feels it is unrealistic to assume that adolescent learners have the responsibility to reflect on a test after it has been returned and its score jotted down. He thinks that in general learners are simply not interested, with the exception of some extremely motivated learners. The "Karin" in the section below is an example of a motivated learner.

I: What do you expect your learners to do with the test when you return it to them?

P: (laughs). Yes, that's quite something. If they're very independent, they'll take that test home with them, [mm] and rewrite it, improve it, and then they'll turn it in once more. And they might come up to me and say, gee, how could I improve this or put this differently? And how could I treat this subject better than I actually did? That's, that's what they should do, but nobody actually does that. I've never seen anyone do it.

I: So you get another fat 'should'. This is what it 'should' ideally be like.

P: What they do is jot down their marks. They check what their overall scores are [yes] and they file the tests because they are required to.

I: No exception to the rule?

P: No exception!

I: So nobody goes up to you and ask you what they should do to?//

P: Well, Karin.

I: Yes, Karin. One of your more motivated learners.

P: Yes, she wants to improve herself. And she actually does.

I: Did I see her today? I believe I did, didn't I? Yes, putative 'Should' is the right word here, isn't it? So apparently there are a lot of 'shoulds' involved when learners are expected to regulate their own learning. [YES (emphasised)] that would //.

P: That, that, that// I find this whole second phase just so (laughter) theoretical. I don't believe that there's a single pupil out there who would crawl behind a pile of books with this kind of weather, do you?

I: I find that hard to imagine, too [No (laughter)]
(3-2-54)

Pete's perceived absence of test washback for the majority of his learners has consequences for how he defines the influence of the test on the ways in which learners learn how to communicate in English, and how it affects their autonomy as learners. In the sections below we will go into Pete's views concerning washback on CLE and LA.

10.4.4 Washback on Communicative language Education (CLE)

As we have seen, Pete does not think the actual tests themselves will affect how learners learn to communicate in English more effectively. He feels the majority of his learners are not interested in potential washback effects. From all of his 4-havo learners, only one female student is interested in the results of the test further than just the marks they get. She regularly asks him for extra materials. Learners tend to be more interested in improving on their maths results.

Yet, Pete had mentioned one important potential washback effect of the *Dear Nobody* test. When we discussed the knowledge and skills the learners were expected to show in the test, Pete was very much interested in 'creativity'(see 3-2-5, 6 & 7 quoted earlier). Learners are encouraged to express themselves in English, and do so with 'creativity', so that he is gripped as a reader. Pete has suggested that creativity is a skill that can be acquired by repeated practice, by answering identification questions such as 'What would you do in a situation like this?' or 'What advice would you give to that person, given the circumstances?'. The result of this practice could be that learners learn to convincingly express their views, thoughts and feelings in English. However, potential washback effects such as these are not exploited by Pete.

10.4.5 Washback on Learner autonomy (LA)

When we explicitly raised the question of the influence of the language test on the development of learner autonomy, Pete mainly focused on how he feels his learners typically prepare for the test. Pete's learners are given a lot of freedom. They are given complete autonomy over test preparation. Of course, Pete gives hints and offers pieces of advice for the learners, but they are relatively free to either follow them up on or disregard them. Pete again suggests that each and every person, including the learners themselves, knows what learners are expected to do. They should study reading texts intensively and extend their vocabulary on the way. They should refresh their knowledge of English grammar, particularly what they ought to know about the English verb, and apply this knowledge to their own writing. They should regularly hand in products in the course of a project. They should consult Pete in case of problems. Pete sadly states that these 'shoulds' generally fail to materialize.

I: How autonomous do you really have to be, to do this test successfully?

P: You really have to be quite autonomous. Because, besides doing things like this, theu should be increasing their vocabulary by critically reading different texts.They should be reviewing their grammar while doing this lesson series. They should, HA (laughter) are you getting all of this? [Yes, just fine] They should be handing in assignments. They should be discussing these things with me. And in all these areas they should be =interruption= And so they

should be doing all kinds of things. [yes] But in reality, none of this is actually being done.

I: You never get anyone who says//.

P: If I don't force them to – and I never really force them, I have to say. I do believe that they think I force them, but as a matter of fact I really don't. I haven't forced anyone to really study that chapter on grammar and then to write a piece and to apply what they just learned and to take another look at it and – oohh.

I: Make sure that all those parts [yes, yes] that you studied are really in it.

P: How did that go again? That //. Basically they all should be doing that, but none of them do.

I: But it's no big deal really, because the final results can always//.

P: Yes, because you adapt, but hasn't it always been like that?

I: Yes. Absolutely. [ha, ha] Yes, responsibility. So that means that there are no learners who willingly and voluntarily do what they are supposed to do?

P: There are a few who do that, but the majority just has to get by on three lessons of English each week.

(3-2-34)

Pete seems to have accepted that the majority of learners only learn by doing the project tasks and activities in preparation to the test. Pete doubts whether his learners get down to practising and studying for English outside the regular lessons. A lot attention and energy has to be paid to the subjects of maths and Dutch in the second phase at his school. Learners have to carefully plan their work and select what to do and what to leave.

P: I don't actually know if they do anything at all for English at home.

I: Yes. I don' know really. I mean, that's one of the questions that was asked//.

P: If they do much of anything at home. At this school, a lot of time is wasted on Math problems.

I: That's what I understood, yes. Also as a result of the first report card meeting, That it's mainly because of maths.

P: The kids here spend a lot of time doing maths problems. And they have absolutely no time left at all to analyse a text. Difficult! They really should. Of course they should. In theory I'm allowed to have so many hours of autonomous working hours, but how do I realise a thing like that? By making them learn four pages of idioms? Well, I just refuse to do that. By making them read another book? The poor dears, (laughter), those little kids. So I do pity them a little as well.

(3-2-35)

Both Pete and Joy had confided they were not too pleased with the idiomatic knowledge of the majority of their 4th formers. That is why we decided to challenge Pete a little by suggesting idiomatic demands of the two reading texts that had been part of the project. Pete does not warm up to our suggestion to specify what learners have to know or should be able to do on the basis of the reading texts. After some

thinking Pete says that by setting demands like these, the learners might no longer enjoy the tasks and activities, and perhaps eventually lose their motivation to learn English.

> *I: But it would be taking things too far, in your opinion, to say, well, there were two texts in this project. I expect you to, well, at the very least, if I ask you to translate eighty percent of this text into Dutch. That you know what it says. Would that be taking things too far?*
>
> *P: Yes, it would.*
>
> *I: Yes. Because?*
>
> *P: - 6s – Because then I believe that English will start to annoy them.*
> (3-2-36)

After a third interruption Pete spontaneously takes up the subject we had been discussing. He stresses once more that having learners learn by setting demands or criteria beforehand is of no use. The learners have their own responsibility. Pete only feels responsible for the learner's active participation in his classroom. Whatever the learners do outside his lessons is not his primary concern.

> *P: Your last question implied coercion. And so why don't I act?*
>
> *I: Yes, it did, in response to your 'learners should, should, they reaslly should, whereas they downright don't..*
>
> *P: Yes, because I really do think that it's their own responsibility. I'm not going to feel bothered when a learner refuses to work. What's the use? I think that you should get that learner to work in a different way. And that's// yes, with a bit of friendly motivation. If they like it// If they think, yuck, I just had maths, now it's English, and another boring lesson. That's even worse for me than what I feel they now think of English, in general.*
>
> *I: So you really go for motivation? [Yes, I do] When learners are motivated, they need no encouragement to get started. [Yes, indeed] Then they actually want to know and understand what's being said.*
>
> *P: And so I think that the responsibility is all theirs. I tell them so often enough, I think. Something like, guys, you should be doing this or that. I think I pretty much compel them to participate in class, but what they do outside the classroom, well, it's their own time, isn't it?*
> (3-2-37)

Pete's views of learner autonomy are also reflected in his class management. If learners do not actively participate in his classes, he simply asks them to leave the classroom. There are no repercussions on Pete's part. He feels it is simply not worth the energy.

> *I: Which means that you do set a certain standard of behaviour in your classes, and so there might be someone, and you're just discussing one of these texts from the project//*
>
> *P: Yes, then I do think that that person should participate, yes.,*
>
> *I: And someone is looking out the window, or chatting and you think, well, wait a minute here.*

P: Yes. Either you're in, or you're out, is what I always say.

I: Yes, exactly. So then that learner goes like: 'Well, I'd better participate.'.

P: Yes, or decide to leave the classroom!

I: Or leave the classroom, I see.

P: Then they really find it odd that I don't do anything. I don't find it important. I'm there because I want to be there. It's what I like to do best, and so in a way that needs to reflect on the learners . It should affect them. And if they don't want to participate, well, I can really imagine that if you come to my class and you've just got a three in maths, that half way through the class you really feel like doing something else. Now, why do I always use maths as an example? But any way, I might just as well have used Dutch.

(3-2-38)

So, Pete feels that his learners are ultimately responsible for how they deal with the subject matter they practise with and the tests they do. He himself feels responsible for creating conditions, such as creating motivating assignments and tasks and offering advice and support whenever and wherever he can. He does so with a sense of realism. He feels there is more to life than school for his adolescent learners. Given these beliefs, Pete did not aim at any explicit positive washback effects of the writing test.

This is where our discussion of Pete's first test ends. The second test we discussed was the *Little Boy* grammar test, which was illustrative of Pete's beliefs concerning language tructure.

10.5 Test 2: The *Little Boy* grammar test

Even though the test was already presented in chapter 8, it has been printed below for ease of reference.

10.5.1 Justification

The second test we discussed with Pete was the *Little Boy* grammar test. Pete said the test was really Joy's choice. He would have preferred to discuss a writing test called *Cal,* a summative writing test of a project on Northern Ireland similar to the *Dear Nobody* test. It again stresses Pete's firm belief in project education. Pete feels grammar is too abstract a subject for his 4-havo learners. Having said this, Pete does not at all mind discussing the grammar test.

P: Yes, well all I can say that this test is Joy's choice (laughter) and that I told her that I would go along with that. I don't mind.

I: For the sake of comparison?

P: Yes, yes, yes.

(3-3-4)

Pete adds a justification of his own. He indicates the test is illustrative of what he feels an effective grammar test should be like. He is convinced it is important for a grammar test to be concrete, realistic and authentic.

P: Well, when you discuss tests such as this one, I find that this grammar test is pretty realistic. What I care about is how useful it is for the kids. And if you

take a newspaper article as a starting point, that's more concrete than all of those isolated little sentences that have been fabricated to really trick children.
(3-3-5)

Pete objects to grammar tests that contain isolated sentences that are full of grammatical pitfalls. He feels such sentences are abstract and rarely ever carry any real meaning. Tests of this kind primarily assess what learners cannot do. Pete feels that the integrative way in which grammar is tested in the *Little Boy* test offers a viable alternative to testing grammar points in isolated sentences.

10.5.2 The knowledge, skills, and insights measured by the test

Pete says the knowledge and skills assessed in the *Little Boy* test are rather limited. Knowledge is restricted to 'knowledge of verb forms' and the skills to 'a little reading comprehension'. According to Pete reading comprehension is involved because context is an important factor in deciding on the appropriate tenses and/or forms.

I: What does this test assess?

P: Knowledge of verb forms. Really no more than that. It's very limited. It does require a little reading comprehension. I think you need the context to be able to fill in the appropriate forms. But that's basically it.

I: Right.

P: So actually it's a rather limited little test.
(3-3-8)

We attempt to learn more about the declarative knowledge involved when a learner is asked to use an appropriate tense and/or form of a given infinitive in a reading text. Pete, however, does not go into any declarative knowledge. Instead, he mentions the procedural knowledge a learner needs in order to arrive at a correct tense or verb form. A learner has to apply the so-called 'when-scheme' to arrive at correct and appropriate tenses and/or forms. The 'when-scheme' is a heuristic scheme that starts with the simple question: 'When does it happen?'. The question is followed by a number of options. To each of the options a tense and/or form has been added.

I: Yes. And when you mention context and verb forms, what exactly do you mean? Do you mean more than knowledge of time adverbials, such as 'so far' and 'now' and//

P: No, no, no, no. Because we're teaching kids according to the 'Pete' system. I keep it very simple. It starts with a sentence. Then comes the paragraph and after that you have to be able to place the paragraph in the complete text. And this test is a result of that. The learners pose the question: 'When does it happen?' Then you get a whole bunch of answers, and if the answer is 'a', then you follow the arrow to where it leads. And that's towards the tense you need to use. I teach them to rigidly stick to the scheme. It's very simple and no more complicated than that. And the test fits in seamlessly with the scheme. They can do the entire test by using that scheme.

I: Yes. And so that scheme implies that you're working with a lot of, let's say, 'time adverbials', such as 'yesterday'. You add a date and//

P: Yes, things like that. Yes, yes.

I: And 'now' means 'at the moment' and so on.

P: Yes.

I: All right. So that scheme is really the tool your learners use. They need to use it to find the correct tense. It helps them to discover the right tense.

P: Yes, it gives them hints, yes, yes.
(3-3-9)

According to Pete procedural knowledge of the 'when-scheme' is a safe option to successfully do the *Little Boy* test. Pete says the learners find the scheme difficult to apply. He wonders why. It surprises him that learners are unable to use such a straightforward scheme.

I: All right, that's clear. So they have to know about the scheme to do well on the test. [Yes] Someone who applies it correctly//

P: And they actually find it very difficult. Some way or other they really find it difficult. And I simply don't get it. I understand perfectly that they find reading comprehension difficult, because that's very vague. But this is very clear-cut. And this little scheme, it doesn't take up more than twenty centimetres of the page, but they still think it's really tough.
(3-3-10)

Later on in the interview Pete offers a reason why grammatical tense and form is such a difficult subject for his learners. He feels many of his junior secondary colleagues for English are to blame. Their beliefs about grammar and the ways in which they teach grammar do not correspond with Pete's.

I: It's intriguing that you use a tight scheme in order to teach your learners how to find the correct tense. So it's a matter of 'If it isn't a, it must be b.' That's almost mathematical. Yet, there are still learners who say they find grammar difficult. They say they can't find the correct tense, even though the correct answer can always be reasoned with mathematical precision.

P: Yes, yes. Yes, I don't get it either. We practise it a lot. And I don't understand that they still can't do this, when they come to me in the fourth form. I don't get it at all. Something must go horribly wrong in the first three years. But we've already discussed this, haven't we. But then that's not just our school's problem. It's the entire country's problem. And I just don't understand it.

I: The question is whether they were ever told. Did they ever practise it? Those are the first questions you pose.

P: Yes, I'm afraid that people just spend an awful lot of time fussing about the passive form, just for the fun of it. And I'm now making an obscene gesture. Yes, it does happen at this school as well, and I ask them why. Why do you spend ten long weeks explaining the passive over and over again every livelong day? You should explain it to me. They say it's very important. And I say why? I don't use it in my fourth forms. And in 5 atheneum they learn to use it on their own. All I want is for them is to be able to recognise it. And, what I'm really getting at, I don't expect my learners to be able to write in perfect English. I find that total nonsense. That's all just relative. Go to any given British high school, grab any given essay and you'll think no way. And that's the work of an actual native speaker.
(3-3-16)

Pete does not refer to his own role as a constructor and educator of the scheme. Pete's statement that learners find the when-scheme difficult to apply does not seem to have resulted in him changing the scheme or adapting his didactic approach. Neither does it appear to have led to any in-depth investigation why filling in the proper tenses and/or forms is so difficult for most of his learners and/or using the scheme seems so hard to do.

The knowledge and skills assessed in tests made by Pete, are never far removed from his pedagogical aims. Pete's teaching is targeted at the learners who need his help most. He considers this his primary aim, next to confronting them with knowledge of the world, life in general and knowledge of themselves (Du: 'levenswijsheden').

> *P: Yes, but I think that those learners who can just pull it out of their hats and don't have to think about it have a knack for it, just like me, and I think you should just let them do their thing, because they're doing just fine. I don't bother about them. That's fine, isn't it? And you'll actually see that they do well in all the areas of the language. That's great. I'm not here to help those learners. I'm here for those learners who think it's all just extremely difficult and who find English one of their more challenging subjects. I have to help them out. That's my responsibility.*
>
> *I: Yes.*
>
> *P: And I also think it's my duty to make sure that they're able to score at least a meagre six on their final exams. I don't care about eights or higher. What I really enjoyed this year was talking to the kids, who came to thank me after their finals and said, gosh, we had been working on reading comprehension for so long, and I did it the way you taught me to and I made a five, not a three. And thank goodness I had a six for the school exam , so I just made the six I needed. And I had about eight learners like that. So, out of fifty, I find that a pretty good score.*
>
> *I: That's definitely true, yes. So that really means that as a teacher, you find it your duty to make sure that people score above that mark. In any case.*
>
> *P: Yes, yes, yes. Yes.*
>
> *I: And you'll do everything in your power to//*
>
> *P: Really that's my only duty. Other than teaching them about life.*
> (3-3-21)

The above has been stated without a trace of irony in his voice. This is what he is after as a teacher. He aims for knowledge of the world, knowledge of life, and personal development. Pete tries to achieve this via project teaching. He attempts to teach his learners how to use English as a tool to communicate on issues that help them develop as conscientious human beings. The learners become better users of English in that process.

Pete has already mentioned that only two areas of English grammar are of real importance to him: use of the tenses and word order (3-3-18). It is, again, proof that the statements he made and the beliefs he expressed in the first interview were not merely rhetorical.

Pete mentions that his way of teaching and testing attempts to develop insight and understanding in his learners. He feels some insight and understanding is created in the *Little Boy* test.

I: In how far do these tests assess insight?

P: Well, you create insight within this limited grammatical area.

I: Does that mean that there are learners who have somehow already appropriated this system either consciously or subconsciously and who can read that//

P: Yes, some learners don't use that scheme at all. They say that they'll just do it the way they have taught themselves or whatever. And when I ask them, well what exactly did you teach yourself, they can't put it into words. But they do well. And others do use the scheme and make a six because of it and others try to use it, but won't make a six because I don't think they're trying hard enough.

I: Yes, but so there are learners who say, well, I'll just fill this in really fast. This doesn't feel right. [Yes] This doesn't look right, so let me change this. [Yes, yes] Sometimes they get it right, sometimes they don't.

P: Yes, they are usually the learners who come up with a host of alternatives to the expected response.

(3-3-19)

We decided to follow up on this issue and asked Pete which of the following two types of learner scoring well he would prefer: the learner using the 'when-scheme' or the learner doing well by simply relying on his/her gut feeling? Pete's answer is in line with his beliefs. He is primarily interested in having the learners transfer knowledge of English tense and form to their writing. Learners who do not consciously use the knowledge and skills offered in the curriculum can neither be called good nor bad. This is because they have not in any way focused on the subject manner that was being taught. Because these learners are successful nevertheless, Pete lets them be. He is more interested in the first type of learner, though. His teaching has mattered to these learners. Pete has managed to add something to what the learner already knew or was able to do. He would like to see himself as the provider of cream to his learners'coffees.

I: In your opinion, which learner is better? Say someone makes a seven on a test and did so by consciously applying the scheme and consequently getting a seven back on his test. Another learner always makes a seven, and doesn't have to think very hard to do so. In your mind, which learner is better, or is it not that black and white?

P: It doesn't really matter to me. The former, who's really studied the scheme and applies it well, if he's able to apply that which he's learned to his own writing, then he's learned more. He's done more to master the English language. The latter hasn't done anything to master the language, which makes him neither good nor bad.

I: Yes. That's just something that person could already do and consequently he//

P: Yes, that's something which is very matter of fact. So I don't even have to wonder whether he's good or bad.

I: So when you speak of learning, it's about becoming aware of what it is you're doing, taking a moment to reflect//

P: Yes, and I find the latter to be of more importance at school. And I find more joy in that. I mean, I'd probably find the other person to be very friendly and

nice, but anyway, he might as well not come to school, because apparently he can already do all those things.

I: You don't believe that you're adding anything to their knowledge of grammar?

P: No, I don't add any cream to their coffees, unfortunately.
(3-3-20)

Pete mentions a crucial aspect here. The ultimate aim of grammar teaching is that creating a better understanding of language form will improve language use. Yet, there are learners who somehow seem to have acquired such understanding. They have always been 'good' at grammar. Some of these learners have rarely ever studied grammar rules consciously. They opt for an appropriate tense or form simply because it feels right. Such learners are often unable to indicate why a certain tense or form is correct. At the other side of the spectrum there are learners who desperately hold on to anything they can learn by heart, such as paradigms or rules. We will discuss the ins and outs of grammar teaching in settings conducive to learner autonomy in more detail in the final chapter of this study.

In the Netherlands, a large proportion of secondary school learners do reasonably well for the subject of English. Average scores of around 7 are rules rather than exceptions. Many of these learners rely on their intuitions and feelings why a particular utterance in English is or is not correct. The same has been the case for his two sons.

I: How big of a role does a person's aptitude for languages play for a student who thinks//

P: It's very big. Both my boys have never really done anything for English and they always get marks around 7,5. Always. It doesn't matter what test they're doing, you know, Tim, 7.3, Tom 7.6, Tim, 7.5. Always.

I: So they do make errors and mistakes, but never fail tests.

P: No, it's really ridiculous. But they don't do a thing, these boys. Well, I guess they got it from their Dad. I don't know. And there are learners who are good at listening, bad at multiple-choice, bad at speaking, fair at writing. So there's some variation there. I find it very tricky to put the marks for the school exams next to the marks for the multiple-choice exam. Because those skills really are different. There are these science-boys, I always call them, brainy boys with horn-rimmed glasses is what we used to call them, they can make a nine on a reading comprehension text, because they have a kind of analytical ability that allows them to do well on that. But you don't want to talk to them. But there are also those who can talk a mile a minute, but who will never understand why answer A is the right one. So it can really differ. I never bother much about all these differences. The only thing might be that their marks on the multiple-choice test are much higher than their marks for the school exams. But it can also be the other way around.
(3-3-27)

Similar to Mark and Joy, Pete is critical of the impact of the national CITO reading examinations the havo and vwo-learners have to take at the end of their secondary school careers. Pete feels that 'a kind of analytical ability' is the dominant quality a learner needs to do well on these tests. Yet, the learners who do well on the CITO reading examinations may be notoriously bad speakers and writers of English. In the course of the year, all of the teachers have mentioned potentially negative impact effects of the national reading examinations on the ways in which learners learn to communicate in English.

10.5.3 The construction and use of the *Little Boy* test.

Construction

As has been presented in chapter 8, the *Little Boy* test has been constructed from an original authentic text. Here we already indicated that Pete constructed the test. In this section we will deal with how he selects and makes tests such as the grammar test under discussion.

Pete selects texts and turns them into verbal-tense-and-form tests in the following way. First, he selects a text that offers a variety of tenses. Then he may add some explicit time adverbials to help the learners use the correct tense more easily. Pete mainly focuses on the variety of tenses, when he selects texts. The actual subject matter is a little less important to him. Only on second thoughts he claims that the text preferably has to be of interest to his learners. The text of the *Little Boy* grammar test had been taken from the Daily Mail, in line with Pete's preference for the use of tabloid texts for this particular purpose.

> *P: I will definitely use a real, authentic newspaper article, check to make sure it uses different verb forms and then adapt a few sentences where needed to make them clearer.*
>
> *I: Yes, all right. But where did this text come from? Do you remember? That//*
>
> *P: The Daily Mirror.*
>
> *I: The Daily Mirror. All right.*
>
> *P: I think, the Daily Mirror. One of the tabloids.*
>
> *I: Yes, yes. So do you pay any attention to the subject? Or is that less important to you.*
>
> *P: No.*
>
> *I: It's more of a question whether it can be used or not.*
>
> *P: Well, although, yes, no, it has to be at least reasonably interesting, to a certain extent, because otherwise it just really won't do. And this one was about a little boy who was ill.*
>
> (3-3-11)

Pete provides a little more detail on how he constructs grammar tests of this kind. First he jots down what variety of forms and tenses a text contains. Next he checks whether all of the tenses and forms have been dealt with in class. It is important that all of the tenses and forms can be retrieved by way of the 'when-scheme'. Then he checks for clarity of the clues that indicate what form or tense has to be used. Finally, he goes for some variety in form and tense.

> *I: But it does mean that you require the text to have a wide variety of time adverbials?*
>
> *P: It should pretty much contain all that I explained on the blackboard. And it shouldn't use too much of the same form in a row, because then they start to think, oh, he's trying to trick me, or something. That they'll think, hey, that's the fifth simple present in a row. That wouldn't be very clear.*
>
> (3-3-24)

However, variety in form and tense does not seem to be characteristic of the *Little Boy* test. The original text shows that the test contains quite a few present perfect tenses. This might have affected the validity of the test. As soon as learners get the first tense and aspect right, other present perfect tenses may follow logically and automatically.

Pete mentions the ways in which he did English grammar tests when he was a secondary school student. He is convinced that looking for a certain logic was already on his mind then, even though he was good at languages in an effortless way. A little to our surprise, Pete claims that a lot of verb forms in authentic texts 'do not fit in' because he feels that English people 'mess about' when it comes to using appropriate tenses and forms.

> *I: How did you tackle tests like these when you were an adolescent?*
>
> *P: I thought it was, well, I was quite good at languages. So with two fingers up my nose, you might say.*
>
> *I: So you look at it and you think, well, actually you didn't have to think.*
>
> *P: No, I didn't.*
>
> *I: You believe//*
>
> *P: Well, that's not entirely true either. [Look] I believe that while I was at secondary school, I was already thinking along the lines of this scheme.*
>
> *I: OK. So you thought, hey, wait just a minute right here.*
>
> *P: I had to find a certain logic to it. And there is. The funny thing is that when you get out the authentic newspaper article, there are actually quite a few verb forms that aren't correct at all. In the end, the English mess about just as much as we do.*
>
> (3-3-13)

In the chapter on Joy, we already mentioned that we managed to retrieve the original text of the *Little Boy* text from the Internet. Having read the text, we had found no instances of inappropriate or incorrect verb forms. We decided to follow up on what Pete meant by his bold statement that the use of tense and/or form is often illogical in authentic texts. We did so by mentioning the disparities between the grammar surveys used at Dutch secondary schools and the written and unwritten rules of English usage. He did not really take up the challenge.

> *I: You could say that. You could also ask yourself what exactly is allowed within the often restrictive rules of school grammar and what in fact is real English.*
>
> *P: Yes, you see, the English imply the rest through their choice of verb form. And I just added the rest to make matters a little easier.*
>
> *I: Could you give me an example? One where you think, hmm, you'd really expect this tense here, whereas the authentic article uses this alternative, which I don't think I, as a teacher, would accept.*
>
> *P: For instance, I think I added the word 'initially'.*
>
> *I: OK. Yes, yes. Oh, that's what you mean. You assume that a number of adverbials//*

P: Yes, just to make things clearer for them. I believe the original was, "The doctors gave him only eighteen months to live."

I: Yes, exactly. Yes.

P: And then I just add a word to make it clearer.

I: So you added an adverbial of time here and there?

P: Yes, I made the hints a bit more obvious.

I: All right, yes. A fine choice.
(3-3-14)

The example Pete gave made us worry a little less. He adds explicit time adverbials to help the learner arrive at an appropriate tense or form and facilitate application of the 'when-scheme'. However, as we have already seen, Pete was not the one who added the word 'initially' to the text. The differences between the original article and the text as it was used in the test was already discussed in the chapter on Joy.

Expertise

Pete's preferred way of grammar testing has a history of its own. His first grammar tests were taken from what he called 'the most progressive course materials available' (3-2-23). Nevertheless, they were still rather traditional grammar tests, characterised by isolated sentences and gap-fill exercises.

P: Yes, we used to do that with sentences and so on and with gap-fill exercises and there were// I think it might also have been the result of laziness. I think about fifteen to twenty years ago, we used to select these course materials, and then we would usually go for the most progressive course materials available, that at least had some coherency to them, but then at a certain point I thought, well, this is insane, because it doesn't make any sense. It doesn't have anything to do with reality. Then what is related to reality? The written word as it exists. But books or novels are out of the question, because that's just going to be too much. So let's have a look at newspapers. Can we manage that? And that obviously leads you to those tabloids, because the language is pretty simple and often fairly juicy as well. And the juicy part often holds the variation, while the sentences aren't too difficult. Well, then I thought, gosh, let me see ,what I can do with this? And from there I got to where I am now.
(3-3-23)

Pete's first colleagues had pointed his attention to more progressive course materials, and to texts that focused on authenticity and meaning. His profound belief in authenticity and meaningful learning made him move from testing discrete grammar points in isolated sentences to main grammar points in coherent texts.

Conditions

The test was administered in a 50-minute lesson. Dictionaries were not used, even though Pete would not have minded if a learner had asked for one.

Preparation

The learners did a number of practice tests before they did the summative informal test. The 'when-scheme' was applied to these practice tests.

P: Yes, yes, yes. I'll go over the 'when scheme' one more time beforehand, before this test. They've actually already learned all about tense over the past three years, but I'll write out that scheme on the blackboard. We'll do some exercises and yes, they should really be able to score at least a seven or an eight. Because of course, there are a few tricky things in there. At a certain moment, when you take an article such as this, you really have to consider the entire article to be able to fill some things in. The aspect of reading comprehension. And yes, that's what makes it a bit tricky. But on the other hand, there are so many forms that they can find easily enough, that it shouldn't be too big of a deal.

I: Yes, yes.

P: But they do find it difficult.

I: Yes, yes. So results differ a lot?

P: Yes, they do.
(3-3-10).

Assessment

When we discussed how exactly he had scored and graded the grammar test, Pete did not become very specific. He claimed he accepted viable alternatives to the tenses and forms that were originally in the authentic text. It leads him to a statement not uncharacteristic of Johan Cruijff, one of the more famous Dutch soccer players: 'All that is not incorrect, is correct'.

I: Another interesting aspect of tests like this is that learners might come up with a certain variety that's different from the solution you have in mind.

P: Yes, that's quite OK, yes.

I: And you think, well, actually I//

P: Yes. All that is not incorrect, is correct.

I: Yes, that's clear, yes.

P: It's not entirely unequivocal, but it doesn't have to be.

I: Certainly, yes. And in that case you think, hey, I expected this response, but then you re-read it and you come to the conclusion that there are indeed alternatives..

P: Exactly, yes. Then when we go over the test together, I say, I find this the most logical answer. But there are some learners who have this or that as their answer and that's OK with me as well.
(3-3-15)

Pete says that his colleagues and he had agreed on accepting alternatives and discussed these if they occurred. Considering his view that native speakers 'mess about a lot' in using appropriate tenses and forms, makes me wonder what alternative tenses and forms Pete had ultimately accepted as alternatives. Unfortunately we did not get down to discussing alternatives he had accepted. Instead, Pete indicated which test items appeared too difficult to do for his learners (3-3-69).

I: Do you remember any items that were difficult for them to answer?

P: I do. 12, 7and 12

I: 7 and 12, I see.

P: Because passives have to be used here.

I: Ah, yes, I see.

We have already discussed that item 7 gives away half of the passive construction that has to be used and that item 12 is not a passive construction. Nevertheless, the constructions, in whichever way they were interpreted by the learners, have allegedly caused difficulties.

P: And for now for something textual. "initially", that must be item 10.

I: I see, yes.

Pete told us he had added the time adverbial 'initially' to the original text to help the learners arrive more easily at the appropriate form, tense or aspect. 'Initially', however, was already in the Daily Mail text. Nevertheless, Pete is right that the time adverbial is useful here for the learners to distinguish between 'initially gave' and 'he has defied them so far' (items 10 & 11). Pete did not go any more deeply into analysing the items as he did.

P: A difficult part was 18, 19.

I: 18 and 19.

P: Apparently they were hard to do.

I: And why 19, because I made a note on that item as well. What was the matter with 19?

P: Because it says: "Why us, what have we done to deserve this?" It has to be turned into a question. I don't think this was seen all of them, but ….

Item 18 seems straightforward. "When we found out" already indicates the simple past tense " we thought" should be used. Recognising the present perfect tense in "What have we done to deserve this?" may indeed have been more complicated. It nevertheless is a distinct resultative perfect. Something that allegedly happened at some indefinite moment(s) in the past ("have we done") has relevance at the present moment of speaking ("deserve this").

The other problem had already been indicated by Pete: the inversion of 'have' and 'we'. It is not clear whether Pete has accepted both "have we done" and "have done" as correct responses.

P: Let me see. What else was difficult. 25 was. And 27, and 29, and that's about it.

I: I see. I'm almost tempted to ask which 29, the first or the last?

P: Yes, number 29. Why do you wonder? Ah, yes, I get your point.

I: Item 29 has been listed twice, but I don't think this will have caused many problems. But you're referring to the final 29? .

P: I am indeed.
(3-3-69)

According to Pete item 25 “Time is running out” went wrong a lot. It nevertheless is a straightforward simple present progressive tense that 4-havo learners should be able to arrive at, either intuitively or by reasoning. If they do not, three and a half years of English grammar teaching on the differences between progressive and non-progressive forms have failed. Pete blames his junior secondary colleagues for this. Mentioning the items that were difficult for the learners has not led Pete to critically evaluating his own didactic approach and procedures.

Item 27 was also mentioned as a difficult item. It would be if a teacher does not accept ‘We try’ as an alternative to ‘We are trying’. Both are acceptable here. As we have seen before, the original text had ‘We try’.

Item 29 has implicit future time reference. Therefore, the verb must indicate this future time reference. Likely solutions are the original ‘won’t happen’ or perhaps its acceptable alternative ‘isn’t going to happen’.

The fact that some irregularities had not been corrected after the first time the test was administered, might be an indication that the test was not adequately evaluated by the teachers who administered the test. Lack of time might be one explanation. A lack of concern an unfortunate second. In all likelihood the learners who were trained in grammatical accuracy would have been bothered or confused by some of the irregularities of the test.

Despite the irregularities, we should not be overtly critical here. The test is an interesting effort to have learners reflect about tense, aspect and form on the basis of an authentic text. The test therefore may well have potential beneficial washback effects if it is critically evaluated by its users, that is by the teachers and the learners alike.

Evaluation

When we discussed the *Dear Nobody* test, Pete had already told us what he feels learners should do with a test once it has been returned. He also realises that few learners will actually do what they should do.

Pete had returned the marked *Little Boy* tests and said that he usually runs through the correct response with his learners. He did not mention any further evaluation details in the interview we had on the *Little Boy* test.

10.5.4 Washback on Communicative language Education (CLE).

Pete is not explicit about any washback effects of the *Little Boy* test on communicative language education. He feels the discrete grammar point that is tested here is too limited to have any serious washback effects on CLE.

Nevertheless, he indirectly refers to a potential washback effect. The two grammar points Pete considers important are use of the tenses and word order. He sees these two as the ‘hammer’ that is needed to construct the ‘house’. Pete’s ‘house’ stands for efficient, effective and correct English language use that enables learners to express what they wish to express. The test tasks may have had a beneficial washback effect in that respect.

Pete’s hammer and house metaphor is embedded in an interview segment in which he expresses three important beliefs. One is related to washback in general and the other two related to washback on CLE. The first belief is that Pete feels it is important to score the correct response instead of scoring the errors or mistakes.

P: Tests are generally not administered to assess what learners can do, but to score what has gone wrong. We do mark the errors, don't we? We don't mark the correct response. Whereas I often tick off the items that have been done correctly.

I: I see what you mean. You think it's more important, to just add the number of correct//

P: Yes, yes, indeed. I do believe in a positive approach, yes. But learners see that quite differently, I think. And what they miss out on, they want the mark and they forget that this particular hammer is needed to build that English house. [Interruption, the fourth one in the course of the interview; a learner comes in, appropriately with a question on a test]

I: Let's take back what we were talking about.

P the lot into the wastepaper basket, but they forget that they only have the hammer, and that now they must try and build the skeleton.
(3-3-25)

Pete says that satisfactory marks are important to most of the learners. He claims that such a preoccupation with marks makes them forget that learning English is a formative process. He acknowledges the power of marks. If learners realise that the results on a particular test is an important indication of how their English language proficiency is progressing, the test seems more likely to have positive washback effects on communicative language education. Pete, however, does not go any further than telling the learners all this and advising them what to do. It is up to them to follow up his advice or to discard it.

Another potential washback effect is hidden in Pete's belief that the meaning of the *Little Boy* text has to be taken into consideration to arrive at an appropriate selection of verbal tenses and forms.

P: Well, yes, in as far as you can call grammar communicative as a separate part, I think that you have to understand the lot a little if you want to be able to fill in the correct response.

I: That's clear, yes. So the meaning of the text and the verbal tenses and forms the learners have to fill in are closely related.

P: I believe they are.
(3-3-73)

As a follow-up question, we asked Pete how he sees his role as a teacher in having the learners distinguish between short-term relief of a good mark and the long-term gain of proficiency in English. The response we get is a mix of a little frustration and the realism that English grammar rarely ever is among the highlights of an adolescent's life.

I: How do you see your role as a teacher here?

P: Well, by constantly harping on the matter.

I: I sense some frustration here. The fact that you constantly have to tell them so. Why is that the case? Why don't the learners pick this up and tell you they//

P: Well, that's not so complicated, I think. I think it is because the weather is beautiful, because they are adolescents, because they have more subjects

than English, fourteen subjects in all, I think, that they have to an AWFUL (emphasis) lot, and because they do not at all fancy doing what they are asked to do.

I: I get your point.

P: 'Sir, can't you do something funny?'
(3-3-25)

Motivation, that is the learners' willingness to engage in doing learning tasks and activities and the subsequent attention they give to the them, was mentioned frequently throughout all of the interviews we had. That is why I asked Pete whether he felt that tests can motivate learners to meticulously prepare for what they are expected to produce in the test. Pete, however, does not believe in beneficial test washback of this kind. Instead, he points at negative washback effects of language tests. He mentions that he is particularly concerned with the learners who always score insufficient marks. He feels beneficial washback only concerns good learners who happened to score badly on a single test. Because they are good learners, they are able to remedy what has gone wrong all on their own. Weaker learners are more likely to feel frustrated.

I: Can a test that is imminent win over learners to put some more effort into getting fair marks?

P: Oh, that is tricky. Very tricky indeed. There is a category of learners who, with every failing mark they get, become more and more discouraged to do anything at all ever again. And I feel it is only the good learners who happened to have scored bad marks, who think, bugger, was I wrong last time, let's see what we can do about it in the next test. But the weaker learners feel embarrassed and discouraged when they get these failing marks, and think, well, whatever I do, it's not helping me at all. But they are also the kids who keep on working without any thinking, who simply cannot handle the 'when scheme', because they feel it is too limited, because in the German lessons the books are filled with grammar, and the French course books are crammed with grammar as well. And now that feller over there is going to write it all onto the board in five minutes, well, I don't believe a thing of what he's telling me.
(3-2-26)

When we first analysed the interview segment above, we wrote two memos. The first memo was on the ways in which a future test was perceived. Whereas the interviewer perceived a future test in terms of positive washback, i.e. as a new opportunity, Pete perceived the test to come as being negatively affected by the previous scores the weaker learners had got. In Pete's eyes, the test to come would only have a positive effect for the learners who had scored below their fair average in the last test they did. In the second memo we expressed our dismay why Pete did not in any way evaluate the effectiveness of his 'when scheme' and his didactic approach in any way, in view of his concern for the weaker learners. Instead, Pete seemed to persist in saying that his heuristic procedure is so simple that any learner should be able to apply it. Observations and interpretations like these, bring us a little closer to discussing and unravelling the complexities of positive and negative test washback.

In the last section on the *Little Boy* test we will present how Pete interpreted washback effects of the project tasks and assignments on learner autonomy.

10.5.5 Washback on Learner autonomy (LA).

We have already seen that Pete does not believe in washback effects of the actual summative test itself. The result is that potential washback effects of the *Little Boy* test are not exploited. However, Pete feels that appeals are made to learner initiative and autonomy in the period leading up to the test.

> *I: Yes. Well, the same question once more, self-regulation and responsibility. Actually, you have said that tests do not affect learner independence. What about this test?*
>
> *P: Yes, that is what I reckon. In this case there may be some effects in the preparatory phase, Sir, I'd like to practise once more; Here's another text; Sir, I'd like to discuss it; Come and see at that particular time. Things like that, but that's all more in general. There's not much going on with the actual test sheet.*
>
> (3-2-74)

Earlier on in the interview, Pete had expressed a number of views related to learner autonomy and washback. He also referred fairly elaborately to the impact of the national CITO reading comprehension examinations, as Joy an Mark had done before him. We are presenting these views of Pete's in the sections below. Even though they do not always directly concern the *Little Boy*, it may add to our reflections and future discussion of test washback in the final chapter. Earlier on in the interview, the matter of how autonomous learners have to be if the only thing they have to do is apply a pre-concocted 'when-scheme' came up spontaneously.

> *I: How Interesting that you raise this. How autonomous do your learners have to be in order to do this test successfully? I mean, you say, here's a scheme, we're in for some practice, we're going to apply that scheme, and we're going to think aloud a couple of times.*
>
> *P: You must have the will and determination to practise. I guess I have about twelve texts and tests like this one. You must have the will to practise at home what we did with the first text in class, and to reason aloud why a particular tense is used for each and every item. And if you don't get it, if you don't say 'when?' and do not answer the question with something like ' at just four years old' and mark the indication of time, meaning it must be that particular tense or I don't know what. If you don't do this, it will go all wrong. At hindsight the kids say: ' Yes, well, if you say so, it does make some sense.' Yes, but come on guys! It does say ' in a few years' time', doesn't it? So there must be some future reference, mustn't there? Yes, and this is where it all ends. They have to come to see this for themselves. In that sense, they have to be autonomous. I keep on telling the kids: 'It's all very nice, but I'm not your teacher here. Because I can't study these words for you. What I can do is putting the 'when scheme' onto the board, but if you all leave it at that! And next, yes, I do sometimes use the image of a carpenter who tries to hammer a nail with his hand, instead of using the tool. 'Yes, so do we, and ouch, it hurts!' And I reply: 'It does, and so does an unnecessary failing mark ! I'd use the tool if I were you'.*
>
> (3-2-13)

Pete's words illustrate the potential of informal tests in affecting what learners learn and how they go about it. Pete also points at the inability of some learners to actually understand what he is trying to do as a teacher. Later on Pete adds some more details on what he expects his learners to pick up from a focus on appropriate tenses and forms.

P: Yes, I'd like to add something myself, if that's all right. This test should be seen as swimming in shallow waters as it were, with a belt around their waists and something to float on in front of them, because frankly it is all very controlled. They're able to apply the scheme here. And they can fill in the appropriate response and arrive at it by reasoning. And, slowly, you have to extend this passive condition to something they start doing themselves. And at some point in time, they will start using 'he hasn't' instead of 'he haven't'. So, the test is a step towards writing proficiency.

I: And you feel this step is very important. It raises awareness. [It does, indeed] They are alerted to//

P: Yes, it does, but there are only a couple of things I need to tell my havo-learners, and to be honest my vwo-learners as well. I think the use of the tenses and word order is quite different from Dutch. Well, I pay little attention to other grammatical areas. In vwo-classes this is gradually extended to full sentences, with clauses and all.

(3-3-17)

By having the learners reason why a particular tense or form has to be used when the learners do a number of practice tests that prepare for a summative test, Pete expects his learners to eventually use appropriate verb phrases automatically and spontaneously when they write in English. Pete feels the 'when-scheme' is just a controlled step towards creating automaticity. He believes that teacher-controlled attention to tense and word order differences between English and Dutch will eventually lead to more learner autonomy and increased language proficiency. He expects a little more of his vwo-learners than of his havo-learners in terms of sentence complexity.

We do not end our section about washback on LA here. In the third interview, with the school year well under way, Pete expressed a number of criticisms that potentially relate to negative washback on LA. He was once again critical of the impact of national reading comprehension examinations and the impact of the second phase reform.

We have seen that Mark had been critical of the second phase reform from the outset. Joy and Pete had been far more neutral towards the year to come. But now, towards the end of the school year, Pete also fell out with what he and his learners had had to put up with in the course of the year. Even though the influence of the reform had not been an explicit question in our follow-up interviews, Pete obviously felt the need to mention its effects.

He spontaneously started his evaluation after I asked him what he felt about the fact that the mark on the national examination reading comprehension test is averaged with one mark that is composed of all of the other marks that the learners have got for all of their school tests and examinations.

I: What do you feel about the relation between the mark a learner gets for the CITO national exam and the marks they have got for their school exams?

P: That has all been determined and decided on for me.

I: A given fact, you simply need to accept?

P: No matter how annoying I feel this really is. When I welcome my learners in the examination classes, I apologise and tell them we apparently had a school year full of fun and cosiness before the summer holidays, but that now those days are over.

I: Because this is what we are heading towards.

P: Indeed, yes.

I: Whether we like it or not.

P: Yes, but I do hate doing this.

I: So next year in havo 5 and atheneum 6 a straightjacket?

P: Well, atheneum 6 won't be too bad, because you still have the old curriculum there, which offers some opportunities, but havo 5 a lot of misery for sure, because, yes, they do have very few lessons and I think I'll have to focus very specifically on speaking skills, listening skills and reading comprehension.

I: From one school examination to another?

P: Yes.

I: Very focused.

P: And mainly focused on closing off without giving anything much attention or thought.

I: I see what you mean.

P: Yes, I think it's awful, but anyway, this the way things are. (3-3-28)

Not unlike what Mark had done in the first interview, Pete now distinguishes between 'the old programme' and the second phase curriculum, which appears to press him and his learners to hurry and haste. Unlike Mark, Pete seems more ready to accept this as a given fact. He tries to cope with it in the best possible ways. Pete says the curriculum will not only be less comprehensive, it will also be less fun for the teacher and the learner alike. Thus, it will affect teaching as well as learning motivation, two of the crucial drives in education.

I: If you compare what you do with your 6 athemeum learners and you used to do with your 5 havo learners with the the second phase in the year to come, what are the//

P: Well, it is much more limited.

I: You're dealing with less.

P: Yes, but that's not all. You deal with less, but the things you do also lack any fun, so that the children do less and less. So, yes, this actually works cumulatively.

I: Learner need to be motivated and engaged to devote themselves to a task (3-3-29)

This is how Pete views his learners' language learning motivation. He censures the learners' focus on short-term gain.

P: Yes, and that motivation does not concern doing a considerable number of reading comprehension texts in November, because they will have to sit their final exams in May. For learners it is all short-term work. Preparation for this test involves six to seven weeks of work. But don't you think that learners

have been working on this for so long. I have, but they haven't. It takes them two days. And that's where it all goes wrong. Because then they start thinking: 'This scheme, what was it about again?'. They haven't internalised it yet. An even if they the scheme next to them, they haven't yet learned to apply it by thinking aloud. It's all short-term work.

(3-3-30)

Pete regrets the learners' short-term focus. He feels his learners do not see the 'when-scheme' as a crucial part of cyclical didactics. Learners tend to brush up their declarative knowledge a short period before the test, without paying attention to procedural knowledge. One of the questions is in how far they are challenged by Pete to go any further, apart from his regular exhortations to do so.

Pete believes that the available contact time he has with his learners does not suffice to learn a foreign language properly. In his case he sees his learners for three lessons a week. In a lot of schools in the Netherlands, 4 havo classes were taught English for four periods a week in the old curriculum. Wherever or whenever the learners learn, Pete is conscious of the fact that he cannot force his learners to learn either at school or in the real world.

P: If they wish to practise, they can do so. But I can't force them to, can I? I see them for three lessons a week, and that's what I have to make do with. And the effective time in which you can get some work done is so much less. They come in. I drivel on about something. And there's always something silly to talk about. So if I manage to get them to work seriously for twenty minutes, I'm more than happy.

I: Are you?

P: Which does not make the remaining 30 minutes any less worthwhile.

(3-3-31)

Pete refers here to the importance of teacher-learner interaction in the classroom. He can truly challenge and motivate his learners. Pete seems keen on stimulating his learners' intrinsic and integrative motivation, and making them feel at home.

Pete feels that the quality of the curriculum is as good as can be expected of a second phase programme. The curriculum Pete and Joy offer, includes the four language skills. Joy had taken good care that the all of the skills had been included in their programmes. The only problem for Pete is that there is too little time to deal with each of the four skills properly. He refers to his second-phase teaching as having become shallower and more cursory than his teaching used to be.

I: In the second phase there are targets to be achieved, which involve a lot of skills. This may affect the quality of the English curriculum.

P: No, not any more. It still includes the four language skills, and we could of course spell out all of the various domains, but that's something Joy already figured out last year, and she says, well, it more or less includes everything it has to include, so that's fine with me. Done and ready. I do not really pay attention to that. Such nonsense. So the skills are still four in number, but the time available has decreased so much that education has become even more cursory.

(3-3-33)

Pete reckons it is not just a matter of contact time being reduced. The reduction was partly caused by the fact that the learners are taught more subjects. Yet, the

main reason for Pete is that many learners are convinced that if they are taught English for two lessons, they only have to study English in these two lessons.

> *I: And by less time available you mean loss of contact time?*
>
> *P: I do, but also the time the learners actually have for studying English.*
>
> *I: Because there are thirteen other subjects//*
>
> *P: Because they have to consider many more subjects. Also because they think like: 'Hey, only one lesson a week for English next year. That's only fifty minutes of work'. It is such a utopian and idealistic line of thought that sixteen-year-olds, and of course I already mentioned this a couple of times, will actually do something of their own accord. Come off it. I wasn't any wiser at that age either, you know.*
>
> *I: Neither was I.*
>
> *P: Of course, you weren't. Otherwise you would have been a proper nerd. (laughs*)
> (3-3-32)

Just as Joy and Mark, Pete has the learners' interests and well-being at heart. Pete tries to identify with the adolescent mind. By doing, he appears to be successful in motivating learners who feel completely helpless and lost with English as a school subject.

We have now arrived at the final test that Pete and I discussed. It is the *Irish Question* reading comprehension test, which he had previously gone into with Joy.

10.6 Test 3: The *Irish Question* reading comprehension test

10.6.1 Justification

The *Irish Question* reading comprehension test selected by Joy as the third test is not a test of Pete's preference. He feels the test contains some bad items. The reason he gives for this is that the first version of the test had been produced by a colleague of Joy's and Pete's. Joy had made a few changes to the original and, shortly before the administration of the test, and Pete had added a number of items he felt were worthwhile including. All in all, Pete considers this a bad test. He will not use the test anymore.

Pete would have preferred discussing another writing test that was based on *Cal,* a novel that was part of Joy's and Pete's northern Irish project. He did not mind too much, though, to discuss this test and complied with Joy's choice for this test. Joy had primarily chosen the reading comprehension test because her 4-atheneum learners had done well.

10.6.2 The knowledge skills and insights measured by the test

Pete finds it hard to indicate the knowledge and skills assessed in the reading comprehension test. Not until he is provoked by the conclusion that reading comprehension cannot be taught, does he arrive at some specification. First he indicates that practice makes perfect. By supplying background information on the subject that is going to be tested, and by mentioning the logical connections between sentences and paragraphs of the texts used for training, Peter hopes that reading comprehension will increase, almost as if by magic.

I: What do the learners have to know to successfully do the test? Reading comprehension?

P: Elusive.

I: Yes, which means learners can't be taught to learn how to read.

P: In a way they can, because you practise with them. You try to make clear how a text has been constructed. You also try to clarify the core of the text. By way of a project, you try to have them read about the subject matter, so that they become familiar with the names that are being used. If you haven't got any clue who Trimble and Adams are, this is all a bubble of air. So in that sense, you somehow teach them how to read. Well, things like this. Vague and indeterminable, I would almost say.

I: But you don't think it's any use to teach your learners to master at least eight words out of the ten words of a given text, so that the meaning if the remaining two idioms van be adequately guessed?

P: No, I don't. 6 and 5 atheneum had a reading comprehension test this year with open questions. A month before the test, I gave them the text and told them to do with it as they saw fit. I told them they would get the actual questions on the text on the day of the test. Then they look at you in an odd way and some start to translate the text, and do not grasp the text at all.

I: I see, so for you there is no relation between knowledge of vocabulary and reading comprehension?

P: No, because the questions are about comprehension. So there's no need to translate the text. A text like this you have to read with the intention to grasp what it is about. You have to see that paragraph one is related to paragraph two, because… And things like that.

(3-4-40)

Pete sees reading comprehension as a skill of recognizing the core or gist of a text and understanding how it has been structured. There is no relation for Pete between knowledge of vocabulary and reading comprehension. We have already learned that vocabulary is no longer explicitly trained by Pete. Alternative approaches to idiom training are not on his mind. Pete simply advises his learners to use dictionaries (3-4-41).

Pete feels that reading comprehension is largely determined by a person's intelligence. He claims the reading skills are more dependent on a person's intellect than they are on mastery of vocabulary.

I: What it comes down to is that learners have to know little of the English language to successfully do a reading comprehension test.

P: It does, indeed.

I: Because reading comprehension primarily requires intelligence?

P: Yes, that's the case with reading comprehension. English reading comprehension has little to do with English, and is more related to intelligence.

I: Do you think that microskills, such as mastery of vocabulary, are far less important?

> *P: Well, we do perhaps transcend havo-level a little, but I once translated a tough vwo-text and the test items into Dutch to help out the kids, but they were making the same errors, because they did not know what the text was actually about. What it concerned.* (3-4-43)

We challenge Pete a little by stating that if learners lack the intelligence, it's impossible to teach them how to read. He feels it is important for the less gifted learners to be self-confident. It is not that hard to score a 5 or a 6 for the national reading comprehension exams. There is no need to get all the items right.

> *P: Yes, no, but also because they have so little self-confidence. I always try to tell them: 'Guys, it's not that complicated. It's easy, really. Listen up. Do pay attention a little.' And then they pay attention all right, but at the same time they think, well he makes it sound easy, but if I have to do it, I won't be able to. And they have to learn to rely on what they actually can do. 'There's no need to know it all, is there? Not all of it, but this particular item is not that difficult, is it?' And if only they could set themselves to using their brains, but often it goes like 'Well, reading comprehension, isn't it?' So they browse through the pages a little, without looking for any cohesion and coherence. If I read a text, it's full of pencil references and notes. This refers to that. Look, this paragraph is referred to over here, or I don't know what, and that section is filled with arrows, underscores, reference words in the sidelines. If learners could set themselves to join in and do things like this, it will in my opinion be possible to have every English bungler score a five. Yes, that's what I always tell them, they are not allowed to write in books, but if you feel like taking notes, go up to the copier and go for it. That's the way in which I construct test items. Then a section may be full of ticks, underscores, arrows and crosses, and I think, well another interesting question coming up. In fact, because I construct these questions, and demonstrate to them what I think when make questions, they should learn how to think for themselves. 'How did Pete go about this again? I see, that's why he made a question like this one.'*
> (3-3-45)

Reading comprehension equals the willingness and ability to look for logical connections between sentences, paragraphs and other prominent sections of a given text. Pete demonstrates and expects his learners to tackle reading comprehension texts in similar ways.

Later in the interview, Pete modifies and partly revokes that reading skills are unrelated to a person's knowledge of English. This is how he returns to the issue.

> *P: Well, in the end I said, yes, that English is not all that important, but of course. Well, it is important to a certain extent, so a learner should somewhat understand the larger part of a text. But I think that goes without saying. Because I do work with English texts in class, I assume they learn to master a fair amount of vocabulary one way or the other. And I think it is important for them to be able to distinguish between matters of primary importance and side issues. And that what I stress in my teaching, instead of paying attention to only recognising the words. But recognising words can still result in a blur, if you don't know what to do next.*
> (3-4-49)

10.6.3 Construction and use of the *Irish Question* test

Construction

The test was initially constructed by a colleague of Joy's and Pete's. Pete thinks it is a horrible test. He mentions two reasons for this: time pressure and the fact that his colleague had included questions Pete found devastatingly simple. Two days

before the administration, Pete added some test items he thought were much better. In the section on assessment we will return to these in more detail.

> *P: I don't know what Joy has told you about this test, but I think on the whole it's a very bad test. Typically the result of time pressure, others come up with the test, who were to construct the test. Not Joy, by the way. And then questions crop up, of which I think, boy, no level at all. What does 'it' mean, and question like that. Yes, I absolutely disapprove of that.*
>
> *I: So questions that were literally about antecedents of words//*
>
> *P: So they come up with this text two days before its administration, kind of like, yes, here are the questions I've got and then additional items have to added in a hurry. But an awful test, really. I won't ever use it again.*
> (3-4-36)

Pete pays a lot of attention to selecting appropriate reading comprehension texts. He feels constructing a good reading comprehension test is largely dependent on the text that has been selected. Nevertheless, it appears to be difficult for Pete to give specific criteria for text selection.

> *P: Well, I have to admit that this particular text was not very suitable for making questions like these. Because you can't just select any random text, which is what my colleague had thought he could do. He thought, I am going to select a text, to which I will then add questions. All wrong. You initially select about a hundred texts, and the eleventh you read may be suitable. And if you get bogged down, you simply throw it away and read the next,and wonder what you can do with that one.*
>
> *I: How do you go about selecting texts? What kind of texts are best suitable for comprehension questions, if you have to find an appropriate one?*
>
> *P: I can't tell you, really. I read a lot. Newspaper articles. And when we're dealing with a particular subject, I start digging in my files. It concerns texts that I had thought would be suitable, and then I start working with these. So I get a pencil and a piece of paper. I ask myself whether I can ask a sensible question about a particular paragraph, which means that a) the learners must be able to answer, and b) which relates to comprehension and not to knowledge of vocabulary. It should never concern vocabulary. It should always concern comprehension. And why I feel that items 7 and 12 are about comprehension, I can't tell you exactly, but that's what I feel.*
> (3-4-37)

Expertise

We asked Pete when he first started thinking about reading comprehension as a construct. It was the moment in his career when he had developed the self-confidence and experience to construct a test that was fully in line with his beliefs. Pete says this happened relatively early in his career.

> *P: Well, the moment you start accepting your responsibilities and tell your colleagues: ' Listen up. I really appreciate using the stuff you have made, but let's now do what I would like to do.' And if those colleagues say: 'Well, OK, show us what you can do', then you commit yourself to the task and wonder; 'Yes, how did they construct the test items and did I like what I saw?' That's the first time you're actually aware of that. That happened in school practice fairly soon. (3-4-55)*

The moment Pete was given responsibility for constructing a reading comprehension test of his own, he wished to live up to that task and started to critically examine both the tests and constructs of his supportive colleagues. Pete says that his academic training had not helped him at all in this respect. As a justification Pete offers the fact that he did his teacher training at a traditional school, while he was already teaching at his present progressive school.

> *P: But don't get me wrong. I was still a student of English and next to my job at this school, I did the school practice for my teacher education at another school. But only came down to forty hours and I immediately fell out with that lady, because I felt she was downright incompetent. So the lady always left when I was teaching. She has never seen what I did. The kids always thought it was good fun, but that was about it. And after forty hours of teaching, I neatly got her signature and I need not hand in a report, because she had known beforehand I would have been rather crushing in my criticism of her. She told me: 'Let's not make such a fuss about it', and I replied: 'That's fine with me.' And that was it. But I was already teaching here in the ways I was used to. And if you then end up with a traditionalist, who had proclaimed as her personal bible a book called Regio, do you remember that? [I do indeed]. Well, I'll be darned.*
>
> *I: Word Study and Regio, and vocabulary course materials like that.*
>
> *P: Awful!*
> (3-4-55)

Both Word Study and Regio had been popular idiom course materials in the 1970s. The ways in which these course materials trained vocabulary were diametrically opposed to Pete's views on effective vocabulary training. Pete advocates:

> *P : Words in context. Look them up, write them down. Do not learn them. Look them up and write them down. Again and again. And at some time or other, you'll arrive at the same idioms, and they are the ones you need to learn, because, apparently, they are the words you need to be able to understand what is being said*
> (3-4-55)

Conditions

Pete has allowed dictionary use. The learners are also allowed to use dictionaries when they do their final examination reading tests. The test was administered in one 50-minute lesson. Joy had only administered this test in her 4-atheneum class. Pete's 4-havo class did an additional multiple-choice reading test the next lesson. Only the test with the open-ended questions is part of this study.

In the segment below, Pete's unorthodox approach to language testing is illustrated once more. He believes he need not be in the classroom while the learners are taking a test.

> *P: Joy only administered this part, but a multiple-choice part was included as well, you know. This text, so the open-ended questions in one lesson, and the next day, say the next lesson, the multiple choice text.*
>
> *I: All right. And dictionaries were allowed?*
>
> *P: Yes, they were.*
>
> *I: Did you invigilate the test or did you just hand out the test papers and leave?*

P: No, when I test I usually am not in my classroom.

I: I'm looking back on the writing test we discussed, where you had them//

P: Yes, I don't think it is that important. I think, but I'm not even completely sure of that myself, that it is important they do the test at school, because then I know for sure that they don't do the test with the help of their uncle or aunt, but I don't care much about anything else. It's impossible to crib or cheat anyway, because the test tasks are pretty diverse.
(3-4-57)

No test invigilation at all is indeed unorthodox, but befits Pete and is in line with some of his views.

Preparation

The "Cal' project on the Irish question had quite a few reading texts. These texts had been studied by asking questions on them that were similar to the questions asked in the summative reading test under discussion.

I: How have the learners been prepared to the test? Did you do a particular number of tests in preparation? Consciously focus on particular texts, or strategies?

P: Well, there were quite a few texts in the project. And we dealt with these by way of analogous questions, of course.

I: You did. So these questions also concerned the, I would not like to call them trivial, but questions that could be answered by literally quoting from the text?

P: Well, no, they were more, let's say, my type of questions.

I: I see. So they were comprehension questions, insight questions.

P: Yes, they were, and once more, being a little pressed for time resulted in a test like this.
3-4-57)

Pete mentions lack of time as the main reason why the test did not have the quality he felt a test should have. Just before the test was administered, he changed a few of the test items and transformed them into his own 'little things'.

Assessment

Our question on the ways in which Pete had scored and graded the test items was preceded by his criticism of the national CITO reading examinations once more. We put forward that open-ended questions had been added to the usual MC-items in the recent examinations. Pete was not overly enthusiastic about the quality of these open-ended questions. He would have liked the inclusion of what he calls 'intelligent questions'. Pete is first of all critical of the fact that the score of the national reading test has a weight of 50% in the graduation mark for modern foreign languages a school leaver gets.

P: Yes, I still believe it's awful. I still believe the [CITO national examination] test is there out of sheer indolence, and I am pretty convinced that it is love of ease. I do understand that reading comprehension is important, because if kids pursue any further studies, they will have to deal with English language reference literature. So reading comprehension is definitely important. But is it important enough to have it make up half of the graduation mark the learners are going to get? Or is it perhaps there because it is the most

convenient and most easily scored instrument to create a kind of equality? I guess the latter is indeed the case. That the national exam is the result of indolence.

I: Indolence and perhaps the convincing reliability of the tests?

P: Yes, yes, indeed. Both aspects, yes.

I: You may be right, there. People have of course started constructing test items that measure more than merely analytical abilities. [Yes] That require a little language production on the part of the learners. Do you welcome this development?

P: Well, yes, I'm not that impressed, you know. Not that I could do it in any better way.

I: What would you like to improve?

P: No, I really can't tell. I don't know. It has of course been a test format that has existed for so long, and which has rendered so many materials that were seen as effective in the past, that I don't usually tamper with it myself. [I see] Although, no, no , no. Well, I do feel you need to ask intelligent questions, and not questions that are self-evident. Can I think of an example from this test?

I: I'm sure you can. (3-4-32)

Pete seemed to or pretended to have missed the irony in the response: 'I'm sure you can'. We were a bit bothered here. There is definitely more to say about the justification of our national reading examinations than 'indolence' or 'love of ease'. The alternative Pete offers are 'intelligent questions'. As examples of such questions he mentions items 7 and 9 from the *Irish Question* reading comprehension test, which have been added to the test by Pete himself.

P: Let me see. Yes, for example in paragraph five, we're now talking about the Northern Ireland test, you see item 7: ' David Trimble says; 'No guns, no government' and Adams says; ' No government, no guns.. What's the difference?' This question goes beyond the actual text and, quite frankly, I think that is a very good test question (laughs). A lot of learners failed to answer this question correctly. They did not grasp what was referred to. But you read, and the actual words are not problematic here. [No, they aren't] Comprehension is what the problem actually is. And this question presents in a nutshell what this text is about. And I would actually call this an intelligent question. I think I'm priding myself now, but what must be must be.

I: I can easily accept that, because you have clearly stated that reading comprehension involves going beyond the sentences of a question or text and that you have to be able to you have to be able what a text is really about.

P: 'Why do Trimble's Unionists object to the word could?' [Item 9]. That one requires similar mettle.
(3-4-34)

After Pete's reference to items 7 and 9 as 'intelligent' questions, he provided more details on assessment. The items score either 5 or 10 points. Pete feels this division makes no sense at all. He particularly objects to the fact that items he considers as crucial, such as items 9 and 10, have only been assigned 5 points. Items he sees as bad, such as 2, 4 and 5, have been assigned 10 points. Pete's

dissatisfaction with the scores for some of the items has not led to any changes in the ways the test was eventually scored by Pete, Joy or their colleague.

> *I: Yes, all right. How have you scored and graded the test? That's interesting, I think, in view of the number of points the learners can get.*
>
> *P: Yes, ridiculous. Downright absurd. I had to arrive at 100 points, I believe.*
>
> *I: Well, let me have a look at the sheet.*
>
> *P: And with an item such as 'Trimble or Adams', item 5 that is, well, yes, a correct gamble will get you 10 points. This does not make any sense at all. On top of this, they only get 5 points for more complicated items, such as 9 or 10.*
>
> *I: But how exactly did you arrive at the division between 5- or 10-point items?*
>
> *P: Well, do count them, together they make 100.*
> *I: Which means the items were not weighted in any way, in the sense that items that better tap into reading comprehension as you see it, will get the learner more points.*
>
> *P: No, because have a look, you can also see that the test has too few questions.*
>
> *I: I see.*
>
> *P: If you want to weight the items in any reasonable way, you will need around 18 or 19 items, because then it would be possible to include differences in points. And in that case, I would also say that item 7 would score 10 points, but for example item 2 would only score 2 points, or something like that. 10 points would also be awarded to item 12, but then again you would need more items to differentiate between them. We haven't got that here.*
> (3-4-59)

Pete believes that the test should have had more test items to be discriminating in an adequate way. We did not discuss the matter any further, even though we failed to see his point why the text would have required 18 or 19 items.

Pete has already expressed his appreciation of items 7 and 9. In addition he also labels items 6, 10, 12, and a little later in the interview also item 8, as effective test items. This means that items 1, 2, 4, and 5 did not meet his standards. Pete did not mention item 11 in the course of the interview.

> *P: There are a number of items that I think are reasonable, but the, the// I think the first question is trash. Second item, pfff, does not make any sense either. Item four is very bad, too, I think. If I give give you the numbers, you will be able to trace them, won't you?*
>
> *I: Yes, I'm sure I will.*
>
> *P: Item 5 I also dislike, because it does not ask after a reason why. It's the question that reads 'David Trimble and Gerry Adams are sent away to negotiate with their followers. Which of the two will, according to Helen Gibson, meet with the more resistance?'. Well, there you say either Trimble or Adams, and this wager may get you 10 points. I'd like to know the reason why, so this item is a bad as well. Item 6 I do like. 'Decommissioning', you won't find that word in the dictionary. [It happens to be in mine]. It may be there, but not in the sense of putting down the weapons and hand in the lot. So, I like that question. Well, item 7 I of course find a terrific question. I like*

item 9, as well. Item 10, too, because the word suddenly relates to what had happened before and they need to have a closer look at that. And item 12 was brilliant as well, and of course you understand that I constructed that item. (laughs)

I: Let's have a look, Ah, yes, I see.

P: But this still leaves us with quite a few questions that actually are neither here nor there. They are about the text, but they are not related to reading comprehension.

I: So, some other time, if you were asked to make this test again, you would only include the items that measure reading comprehension as you see it?

P: I would. I would, indeed.
(3-4-36)

I asked Pete whether he had assigned scores lower than the maximum points to be had for each of the test items. When he said he had, I asked for some details. Unfortunately, he was not able to give me any. When I asked for details, Pete went looking for the tests, only to discover they had already been returned to the learners.

Evaluation

Pete tells me there was no time to discuss the test after it had been returned. None of his learners seemed to have minded this. Pete feels that the scores his learners got was the information they valued most.

I: OK., fine. The return of the test. Its discussion. Are there any particulars in this area?

P: Yes, there is one. Lack of time.

I: Lack of time?

P: We did not actually get down to that.

I: I see. It was returned in the one of the last school weeks?

P: I don't know for sure. No, some time before.

I: No, hang on, of course it must have been returned somewhere in May.

P: Yes, that was another of those messy periods when lessons were cancelled, and sometimes you see your classes only once a week, which makes it impossible to look at a test like this in any detail. Because a lot other things have to be taken care of. No, time, time, time.

I: And your learners do not mind this? They hear their marks, and//

P: It doesn't interest them at all. I haven't heard anything about it. Maybe, it's because of me, but ..

I: Not even motivated learners who tell you 'Please listen do this, Sir' [None at all] 'I'd like you to discuss this.' Hasn't happened. OK

P: But that's the usual response I get, you know. Whatever I return. I have//, yes, you know, if it's a,b,c and I have checked a little carelessly, I may have crossed one single 'a' too many. I always run through the tests with them this way. And sometimes someone says: 'You said 'a' , and that's what I have, so

it must be right.' Well, of course I'll put that right. But in general when I return school examinations, writing skills, literature or whatever, they're always allowed to harp on about something. I tell them to. You can always nag, but you rarely ever succeed. But you can always try.

I: I see.

P: But no one ever shows up.
(3-4-63)

The *Irish Question* test was neither discussed nor evaluated with the learners. As a first reason Pete offered the pressures of the second phase reform that resulted in lack of time. Having said that, even if he had had the time, Pete wonders whether an evaluation of the test would have done much good. Pete states his learners are only interested in product evaluation at the most. The learners are keen on checking Pete's arithmetic and looking for opportunities to move up their scores a little. Washback does not seem to be on the learners' minds. Neither is it a concern of Pete's.

10.6.4 Washback on CLE

I asked Pete in how far the *Irish Question* test challenges learners to improve their communicative skills in English. Pete, once more, expresses his disbelief in the potential washback effects of a single test. It is the project tasks and activities that matter. Pete would just as well do without any tests. He feels that in such a case his assessment of a person's communicative skills would be just as reliable and valid. (3-4-64)

It is once more shown that Pete sees tests as summative assessments. Tests are simply procedures to check whether practice by way of project tasks or activities has led to a particular score in a given area. As such, the test do not play any role in formative evaluation.

Interestingly, Pete in this interview said that he believed that writing tests such as the *Dear Nobody* test more effectively relate to communication. In such tests the learner is challenged to really communicate in a way that is close to real and authentic communication. Such a link is missing in the *Irish Question* test.

I: Yes, sure. So you're saying that a test is a certain assessment procedure at some time, which does not reveal any information on the quality of learning. Neither does the test lead to a particular way of learning.

P: I agree, that is to say, with tests like these. It's different when writing skills are assessed. Because the learners tend to reveal more of themselves than just// and, of course, I do expect them to. I always tell them: ' I have to read those things. In a way, you have to please me, and if you limit yourself to what has to be done, if you can only think of perfunctory blabber, I will just fail to appreciate its content. And they do know how I feel about this.
(3-4-67)

Pete's seems to see communicative skills basically as written production skills in which English is successfully used as a tool for self-expression, communication and development. He now suggests implicit washback effects on CLE of the *Dear Nobody* test, which he did not make explicit when we discussed that test. When we discuss washback in our final chapter, we will go into the roles of implicit and explicit washback effects in foreign language education.

10.6.5 Washback on LA

As has been expressed before, Pete only believes in potential washback effects of the tasks and activities the learners carry out during preparation and training. These preparatory tasks may be potentially beneficial according to Pete. Potentially, but rarely ever in reality. This is because he says his learners often lack the willingness, attention, focus and stamina to relate the tasks and activities to a test they are preparing for or to improving i.e. their reading comprehension as the ultimate aim. This is all accepted as a fact of life by Pete. He believes you cannot force learners into learning. The only thing you can do is to attempt to motivate them. Besides, the context of the second phase reform is seen as more disturbing than was earlier envisaged. Pete and his learners are confronted with the pressures of the new curriculum. Science subjects such as arithmetic, chemistry and physics and new subsidiary subjects such as Cultural and Societal Education (CKV-Cultureel-maatschappelijke vorming) take up the bulk of his learners' time. Pete is very much aware that his learners are confronted with no less than fourteen subjects in the revised curriculum. Both the fact that you cannot force learners into learning and the curricular pressures of the second phase affect the ways in which his learners learn.

This is how Pete responded to my question how autonomous learners should be when they prepare for the reading comprehension test:

> *P: Well, I'm not too sure about whether this test requires a lot of learner independence. The actual learning does, because I am convinced that by repeated practice, you will get better. And that's where it often goes wrong. Kids know they're going to get a reading comprehension test in week 10, but they do not realise, even though their English master has told them a thousand times, they should actually be concerned with the learning process from week 0 up to week 10. What they do, if they do a thing at all, is realise in week nine: 'Shoot, a test coming up next week. Let me see. Shall I do a little text? Yes, let's do one.' And what they conclude depends on the results of doing this preparatory text. If they do badly, they will feel more anxious and stressed for the actual test that is to come. It has a negative effect then. If they do all right, nothing happens at all. At the very best, they may feel a little more relaxed for the actual test. But in both cases, it is not related to a learning process. And that's what I already said in the first interview, or in the second, what they all should do and what they actually do. There's a huge rift between the two. But I'm not the only one who is after them, aren't I? All those math problems they have to be concerned with, and the like.*
>
> (3-4-50)

Feedback has been identified as an important parameter of learner autonomy. It may also be a crucial factor in creating positive washback of the tasks or assignments the learners carry out, either as practice or in summative tests. Concerning feedback, Pete stated some problems. It all started with his view that he feels the English lessons do not suffice to build up enough knowledge and skills to successfully learn how to communicate in English. He feels extra input, practice and feedback are required. According to Pete, computers and computerised feedback might be of use here. In the case of personal feedback, Pete observes how wonderful it would be if he was allowed to individually work with and monitor a single learner. However, he feels that school practice does not allow for such differentiation.

> *P: And what they could do is working with computers. Whatever you feel about this, if a question is done incorrectly, the computer tells you so and may even add what you have done incorrectly. Actually, I should be able to cut myself up in 24 parts and work with a single learner for 24 hours. Even if I could do that for only one hour, I would be more than happy. Classroom practice is so much different. Imagine twenty learners. Five of them have done item 1 incorrectly. The others did not experience any problems. This means that 15*

> *learners will cause noise of some kind. And the two learners who are really interested in the correct response are so distracted because of the noise, they won't be able to follow me any longer. That's everyday classroom practice.* (3-4-51)

Pete states the problem that it is hard to teach learners to effectively communicate in English if you only see them for two or three lessons a week in classes that often consist of over thirty adolescents, who all feel there is so much more to life than attending school. Pete says that the larger the number of learners is, the more difficult it is to individualise feedback. He feels that one-to-one feedback, even if it is for only one hour, would be very effective. Ambitious targets for communicative foreign language skills and increasing learner autonomy can only be realised if the learners are more often immersed in the English language and receive more feedback at the same time. It is an issue we will take up in our final chapter.

This finishes off our discussion of Pete's tests. In two final sections, we will again discuss in how far Pete's core beliefs are reflected in his assessment and evaluation practice. First, we will highlight some general characteristics of Pete's assessment and evaluation practice. Next, we will more specifically deal with Pete's beliefs in relation to the analyses of the tests and the ensuing interviews on them.

10.7 Pete's core beliefs in relation to his assessment and evaluation practice

Evidence of Pete's core beliefs in his language tests appeared to be converging in a way similar to what he have reported on in the chapters on Joy and Mark. This chapter on Pete also provides predominantly converging evidence of how a teacher's core beliefs are reflected in his/her assessment and evaluation practice. We will first mention some key characteristics of Pete's testing practice. In the section that follows, we will more specifically discuss each of his core beliefs in relation to the analyses of the tests and the ensuing interviews on them.

10.7.1 General characteristics

All of the three tests Pete has selected provide convergent evidence of his core beliefs in the sense that:

- two of the tests are based on and are strongly related to thematic projects that include a variety of communicative individual and group tasks;
- the only test in which grammar was tested focused on verbal tense and form and was constructed by changing the verbal tenses and/or forms of an authentic text into infinitives. Test preparation, again, involved a teacher introduction, individual work and group work.
- Pete prefers developing and constructing his own materials to using regular course materials and/or the tests that cone with them. He feels the traditional course materials too often contain tasks and tests that fail to be interesting and lack any real meaning. As an alternative, he develops projects on themes the learners can relate to. These projects are characterised by a variety of tasks and activities in which general knowledge, linguistic knowledge, and language-related skills are all integrated. The first and third test he had selected for discussion were summative tests at the end of a project. Despite the fact the two tests seem to focus on one language skill, i.e. writing or reading, the tests seem to require the integration of general and subject-specific knowledge and skills that were built up in the course of the various

tasks and activities the learners were asked to carry out. Pete truly proved himself to be a project man.

In the next section, we will review the twenty core beliefs of Pete's we illustrated in chapter 7 and summarised at the beginning of this chapter. Converging evidence was found of eighteen of Pete's core beliefs. No evidence was found of two of his beliefs. We will first discuss the beliefs that more or less prominently featured in Pete's assessment and evaluation practice as it was monitored in this investigation.

10.7.2 Core beliefs in the tests and interview data

1. developing his own materials and/or selecting from authentic materials that are available as viable alternatives to the text- and/or workbooks used in secondary education. He feels the regular course materials do not offer what he wishes to transfer to the learners in his lessons;

   A core belief of Pete's that prominently features in his assessment and evaluation practice. Pete constructed the *Dear Nobody* writing test and the *Little Boy* grammar test, all in close collaboration with Joy. As an early and frequent user of the audiovisual technology and the Internet, Pete managed to retrieve a host of appealing texts, tasks and activities, which he subsequently adapts and extends to his own needs. All of the three tests were based on authentic texts. The *Irish Question* reading comprehension test was not constructed by Pete, but was adapted by him on short notice. Irrespective of the quality of this test, the text linked up well with the project it was to close off.
   Pete and Joy started working with a set of regular course materials called *Touchdown* at the beginning of the school year. This set was soon abandoned by the two respondent teachers in the course of the year, perhaps with the exception of some useful background grammar, hints for the language skills and hints to self-assess English writing. We did not discuss the regular course materials in detail. Neither did we talk about the exact reasons why *Touchdown* was abandoned. This was because the tests Joy and Pete had selected for discussion did not or hardly ever relate to the course materials. Pete, as well as Joy, appear to have preferred their own projects to the texts, tasks and activities offered in the course materials.

2. the value of collegial consultation and cooperation from the very start of his career onwards;

   Thanks to the progressive colleagues who educated him as a beginning teacher, Mark believes in working together with fellow teachers who are willing to share and improve his teaching tasks and tests. He consults with Joy on all matters that concern teaching and testing. Pete enjoys Joy's constructive criticism, her attempts to constantly wishing to improve her assessment and evaluation practice and her willingness to discuss educational matters. Whereas Pete is a great improviser, who is not always that precise, Joy is a meticulous and precise teacher. Pete knows he can rely on these qualities of Joy's, e.g. when a PTA has to be written. A fortunate unison of two different, yet kindred spirits who get so much more out of education because they discuss and cooperate.

3. progressive change, which involves critically discussing and challenging educational practices;

Pete was fairly open-minded to the second phase at the beginning of the year.
He did not expect a lot would change in his didactic approach, which he believed to fit in well with what was expected of him in the renewed second phase. To be on the safe side, they bought a set of regular course materials to use with their classes if need be. Yet, these materials were soon abandoned because, once more, they did not offer what the materials and projects did that he was used to developing. Pete critically monitors whether changes are changes for the better. This was not always the case.

4. being given a high degree of autonomy by his school management, which gives him the opportunity to put into practice his particular approach to teaching English;

   Pete respects his headmaster, and his headmaster appreciates him. Pete has not referred to any changes in his autonomy as a teacher in the course of the year of data collection. Largely, he was allowed to continue doing what he had been doing over the years.

5. presenting his learners with thematic projects, in which an array of knowledge and skills have been integrated;

   A belief of Pete's that is convincingly illustrated by the discussion of all of the three tests, of which two had been directly linked to a project and of which the remaining test illustrated how he felt linguistic knowledge should be viewed, tested and integrated in other skills.

6. being critical of the CITO examinations that are used nationwide for reading comprehension, listening comprehension and writing skills;

   Pete still feels that the skills needed to do these tests are rather limited. As an illustration he has offered the help he gave to one of his sons, who was doing badly on his French reading comprehension test. Within two weeks of Pete's help, his son managed to score 7s and more.
   As far as the Cito listening tests and writing tests were concerned, there was only partial evidence by way of the listening and writing tasks and activities that Pete used in his projects, which differed considerably from the format of the listening and writing tests developed by CITO in the year of data collection.

7. being learner-oriented, which requires teachers to engage their learners and partly adapt their teaching to the wishes, knowledge and skills of the learners they teach;

   There is plenty of evidence that Pete really acknowledges the positions and developmental stages his learners are in. He might have been helped here a little by his two sons, one son who had recently graduated and the other son who was about to finish his vwo education. He knows that there is more to life for adolescents than going to school. He also knows that in order to do a little work, learners have to be engaged. That is why he attempts to construct projects that are just as meaningful for his learners as they are for him. The bottom score of 5 for the *Dear Nobody* test and the importance he attaches to his learners feeling relaxed are also examples of his learner-orientation. All without a shade of teacher rhetoric.

8. the fact that the result of tests are important to his learners, even though he is far from convinced of the use of language tests himself and often refers to their limitations;

   Unlike all of the integrated tasks and activities of his projects, the summative tests generally assess only a limited part of what has been learned in the course of a project. Nevertheless he tests, because that is basically what is expected of him by his learners, school management and by the parents. As long as the tests are meaningful and he is entertained and engaged while scoring and grading the tests, Pete does not really mind.

9. Carefully and responsibly preparing his learners for the tests they have to take;

   Test preparation has always been elaborate in all of the tests we discussed with Pete. Test preparation is a responsibility he always takes seriously, sensing that it also takes away some pressure from the young shoulders of his adolescent learners.

10. The limitations of giving marks that are ‘objective’ when complex skills such as writing are being tested;

    It has proved to be important for Pete to be fair and objective in his assessment of writing skills. Painstakingly putting the writing tests into various piles after he has first corrected the tests in pencil, illustrates how important it is for Pete to be as objective as he can possibly be. It is also important for him that the learners are convinced by his sincerity and efforts to truly be objective. A difficulty for Pete in assessing complex skills such as writing skills is that he cannot indicate in exact terms what knowledge, skills and understandings he is assessing when he scores and grades the tests.

11. holistic teacher assessment of complex learner skills without rigorously defined assessment criteria;

    His inability to clearly define the knowledge, skills and understandings that are assessed in all of the three test, results in assessment procedures which are for a large part intuitive and holistic. This even concerns the *Little Boy* grammar test, which was practically a discrete-point test, for which, possibly, some microskills related to reading comprehension were required. The follow-up interviews clearly helped us to gain more insight into when and why a teacher might turn to holistic assessment.

12. integrative testing, based on meaningful input by way of appealing texts or video materials;

    The fact that Pete, similar to Joy, prefers integrative tests, because they so much better approach real-life tasks and activities, is illustrated by all of the three tests. It even goes for the *Little Boy* grammar test, in which he seems to strongly invite the learners to relate cut-and-dry rules to authentic texts.

13. tests that ask for more than the mere reproduction of e.g. grammar rules or idioms without any meaningful context;

Similar to the comments we have made before, it is important for Pete that learners learn to use and apply discrete knowledge and skills in real communication and relate them to authentic and realistic tasks and texts.

14. carefully planning the tasks, activities and tests over the school year;

    Pete is used to planning the tests before deciding on the curriculum. He is greatly helped by Joy here, who is precise and willing to draw up a PTA that is according to the regulations. All of the tasks, activities and tests have been planned before and were practically ready for use at the beginning of the school year.

15. engaging the learners when dealing with literary texts in a light and relaxed way. Pete objects to the artificiality of mainstream literature teaching in secondary education. (3-1-24);

    Literary texts are almost always components of the projects Pete constructs, as has been illustrated by the *Dear Nobody* project and the *Cal* project he mentioned when we discussed the *Irish question* reading comprehension test. Pete takes care to keep the literary extracts relatively short, and to select the parts that are humorous and do not require knowledge or skills the learners are not expected to have mastered. Theoretical literary notions do not seem to be on his mind at all. Literary texts should first of all entertain, with the fact that something is to be learned from interpreting them as a boon.

16. the fact that many learners consider reading novels in English as a real struggle;

    Engaging extracts from novels are presented as ‘appetisers’ in his projects. This is particularly useful for his havo-learners, who seem a little less interested in reading than vwo-learners. Pete is not critical at all of the fact that learners only have to read three novels in English before they graduate.

17. the need to bridge the differences in educational approaches between the teachers at his school who teach forms 1 to 3 and Joy and himself as upper-level teachers;

    In preparing his learners for the *Little Boy* grammar test, he specifically pays attention to the ways in which the learners have so far been dealing with grammar. By having them focus the essential grammar point of tense and verbal form, he hopes to achieve that learners transfer this knowledge to their writing and speech. Nevertheless, he is a little frustrated by the number of learners who still do not understand what English tense and form is all about.

18. approaching the innovations of secondary education in a realistic and positive way;

    Unlike Mark, Pete had not been downright negative about the second phase. Nevertheless, he signalled the problems the learners had to face from the very beginning, and tried to adapt his programme accordingly. In the course of the year of data collection, however, it became clear that Pete does not really believe in learner autonomy for adolescent learners. You can all teach them what they ‘should’ do, but most of the learners simply do not act accordingly. Therefore, Pete fully accepts his responsibility as a teacher in

trying to ensure that his learners keep on building the knowledge and skills they need to communicate in English, orally and in writing.

Eighteen of Pete's core beliefs were illustrated in his assessment and evaluation practice. No evidence was found of two of his beliefs. We present the two below, before turning to a summary of the present chapter.

- the fact that his didactic approach prepares his learners well for the national examinations of reading and listening comprehension;

  In the year of data collection, this belief of Pete's was not directly raised by him. He still seems to think that national reading examinations require specific skills, as well as knowledge about and experience with the test. Pete did not at all mention his learners' abilities on listening comprehension.

- the limitations of a so-called 'native-speaker level' that Dutch teachers of English are expected to aim for;

  Even though Pete expressed this belief of his in no uncertain terms at the beginning of the school year, it did not explicitly return when we discussed the three tests. We would like to argue that this was largely because the tests did not aim to assess a construct such as 'native-speaker' fluency. However, this belief of his might be indirectly related to his difficulty in being precise about exactly what knowledge, skills or insight is measured in a test.

10.8 Summary

In this chapter we started with a recapitulation of Pete's core beliefs and construct definitions as expressed in the first interview. We then focused on the three tests that Pete had selected for discussion in the course of the school year. It appeared that Pete preferred to do the interviews on the same tests that Joy had selected for discussion. He felt Joy's choice was as good as any and he felt it would be interesting to further discuss his experiences with the tests with his colleague Joy. Therefore the tests we discussed were respectively the *Dear Nobody* writing test, the *Little Boy* grammar test, and the *Irish Question* reading comprehension test. The *Dear Nobody* writing test was a summative achievement test on what had been learned in the course of a two-month project on teenage pregnancies. The project itself was based on Berlie Doherty's novel *Dear Nobody* and also used extracts from the filmed version of Roddy Doyle's *The Snapper*. The second test we discussed was a grammar test that had been administered some six weeks after the start of the school year. It was based on an authentic text taken from the *Daily Mail* , which was entitled *Little Boy Growing old before his Time*. The text had been changed into a gapfill text by replacing thirty verbal tenses and/or forms by their infinitives. The learners had to replace the original tenses and forms with the help of implicit and explicit information in the text. The final test Pete reflected on was the *Irish question* reading comprehension test. It again was a summative achievement test that concluded a project. The test had been constructed by a colleague of Joy's and Pete's, with some changes added by Pete. Despite the fact that the tests had initially been selected by Joy, they also provided evidence of eighteen of Pete's core beliefs as they were illustrated in chapter 7.

The three tests Pete focused on were again discussed in terms of *justification*, the *knowledge, skills and/or insights* the tests were supposed to assess, details on their *construction and use*, that is information on how the test was constructed, the

knowledge and skills the teacher required to construct the test, the ways in which the learners had been prepared to the test, the test conditions, details on the ways in which the test was scored and graded, and finally the teacher's expectations of what learners (should) do with the test after it has been returned. The final two aspects that were discussed were w*ashback* on *Communicative language Education (CLE)* and *washback on Learner autonomy (LA),* in which Mark explained in how far the preparation, administration and discussion of the test stimulate the learners to learn more independently and responsibly.

In the next chapter, the data from chapters 8, 9 and 10 will be summarised, put into tables, and compared and contrasted with the data from the theoretical chapters on learner autonomy, communicative language education and foreign language assessment and evaluation.

CHAPTER 11: CROSS-CASE ANALYSES

11.1 Introduction

In three consecutive parts, we will reduce, compare and contrast the data of *Testing for Autonomy* and relate our findings to the constructs of *learner autonomy, communicative language education*, and *foreign language assessment and evaluation* as they have been discussed in chapters 3, 4 and 5.

Our discussion will focus on the similarities between the respondents, without losing track of the differences in interpretation. As mentioned in chapters 8, 9 and 10, the differences in interpretation were shown to be related to their core beliefs about effective teaching and testing.

Each of the constructs will be discussed in four sections. We will first of all present the respondents' *interpretations* of the construct under discussion, as they have been illustrated in chapter 7 *Three Stories to Tell*. Second, we will refer to similarities in the respondents' construct interpretations. In a third section, we will consider in how far evidence converged of the respondents' initial interpretations of a construct in the analyses and discussions of the three sample tests the teachers had selected in the course of the school year. Finally, we will compare and contrast our findings with essential parts of the theoretical chapters on the constructs.

The chapter will provide the input for our discussion chapter entitled *Autonomy tested*, in which we will answer the research questions of this study, using and interpreting the data presented in this chapter.

11.2 Learner autonomy

The three respondents approached the construct of LA from three different perspectives. In the first section, we will argue that the three perspectives are not mutually exclusive. Combining the three perspectives is likely to lead to a better understanding of the complexity of the construct of learner autonomy and some of its practical applications.

In addition, we will also highlight four similarities between the three respondents. The correspondences have been identified as *foundational teacher initiation and control*, *gradual release of teacher control*, *full learner responsibility not being a feasible construct*, and *learning by doing*.

11.2.1 The respondents' interpretations of LA

The three respondents interpret learner autonomy from different perspectives. Joy highlights learner reflection, Mark focuses on knowledge-based competences, and Pete refers to aspects related to didactics. All of the three interpretations are valid in the sense that our discussion of autonomy in chapter 3 has shown that learner autonomy is a complex construct that is multi-facetted and can be interpreted and approached from a variety of perspectives. This means that the foci on learner reflection, knowledge-based competences, and on didactics may all be effective in fostering autonomy in the adolescent learner who is learning to communicate in English in a formal educational setting.

Table 11.1 summarises the teachers' construct definitions as they were offered in the first interviews at the beginning of the school year.

JOY	MARK	PETE
LA involves learner reflection on:	**LA involves three competences:**	**LA involves didactics geared at:**
-what they are doing; -what they have to do; -what they want to do; -what is required to achieve this.	**-planning skills; -self-assessment and evaluation skills; - ability to balance learning and doing.**	**-gradual increase of task and test complexity; -the learner's willingness to learn by doing; -forms of peer learning.**
Role of the teacher:	**Role of the teacher:**	**Role of the teacher:**
-gradual release of teacher control, related to the belief that full learner responsibility is hard to achieve.	**-initially strong teacher control, related to the belief that full learner responsibility is a step beyond self-regulated learning.**	**-gradual release of teacher control, related to the belief that full learner responsibility is a vague concept, beyond the reach of fourth-formers.**

Table 11.1 The respondents' interpretations of learner autonomy

We will now explore the interpretations above in more detail and argue that the three perspectives may complement one another and, therefore, deserve to be united.

Joy stresses the ability and willingness of learners to *reflect* on the present and future tasks that are set for the learners or the learners set for themselves. In addition, Joy wishes her learners to reflect on the knowledge and skills required to successfully do what they have to do and what they want to do in English. Joy's distinction between what learners are expected to do, because of curricular demands or because the teacher simply tells them to, and what the learners themselves would like to do implicitly refers to motivational aspects. In chapter 3 we referred to motivation as a complex construct with extrinsic, intrinsic, instrumental and integrative characteristics and discussed its links with LA.

Mark interprets LA in terms of definable *competences*. The learners should first of all learn how to plan and, more notably, how to learn to stick to their planning. In addition, they should learn how to monitor themselves, which in Mark's case is often related to a review of knowledge components such as grammar, idioms, or the literary notions he feels are required to analyse and interpret short stories or novels. The self-assessment and evaluation skills that Mark mentions, parallel Joy's focus on reflection. However, in Mark's case the assessment and evaluation skills are more closely related to aspects of knowledge. In Mark's perception, the required planning and monitoring skills result in the learners' ability to balance learning activities and doing tasks. Mark feels knowledge is both a necessary and a definable component of skills.

After Joy's and Mark's foci on reflection and competences, Pete adds yet another perspective on LA. He feels that learner autonomy is likely to increase if three principles are anchored in a teacher's didactics. A first principle is that the teacher has to ensure that the complexity of tasks and tests gradually increases. The gradual extension should be implemented in such a way that the tasks and activities will always be well within the learners' grasp. A second principle of Pete's LA didactics is that learners have to be put into positions where they are ready to accept that English is learned by doing and that they develop the confidence to overcome any inhibitions to use the language productively. A final didactic principle of Pete's geared at fostering autonomy is that forms of peer learning are required to develop autonomy, such as peer help and peer negotiation.

We would like to argue that the three perspectives mentioned by the respondents are far from mutually exclusive. It would be perfectly possible to have an educational

perspective on learner autonomy that includes a focus on learner reflection, on knowledge-based competences and on the particular ways in which learner autonomy can be implemented in the teaching and learning process. The argument paves the way for a plea to adopt a *broad* and *deep*, rather than a *narrow* interpretation of LA in formal educational practice. Such a broad and deep interpretation is driven by theory as well as inspired by classroom research.

So far we have been focusing on the differences in the ways in which the teachers have specified LA. In the section to come, we will highlight four similarities.

11.2.2 Similarities in the respondents' interpretations of LA

Even though the respondents interpret learner autonomy from different angles, four similarities come to the fore. A first similarity is that all of the three respondents take initiative and control at the beginning of the learning process. We will discuss this aspect as foundational teacher initiative and control. A second likeness between the three respondents is that they all believe that teacher control should in the end be released, but that such handover of control is a gradual and often non-linear process. This second correspondence will be discussed as gradual release of teacher control. A third similarity we will discuss is related to the ones we have just mentioned. The respondents do not believe in full learner responsibility as a feasible concept in upper secondary education. A final prominent similarity is that all of the teachers believe that learner autonomy relates to tasks and activities the learners have to do, and which therefore can be observed. This correspondence will be referred to as learning by doing.

Foundational teacher initiative and control

Evidence of this aspect of learner autonomy was soon converging in the year of data collection. All of the respondents organise and initiate what the learners are supposed to learn and how they are meant to go about it at the beginning of the school year. We have already seen that the tasks and activities the learners are asked to do, are strongly linked to the respondents' beliefs related to effective teaching and testing.

Thus, Joy and Pete first attempt to bridge the gap they feel to exist between the almost exclusive focus on form in the first three years of secondary education and the focus on communication they strive for. They do so by first of all limiting the grammatical focus to tense and verbal form and secondly by having the learners transfer their knowledge of verbal tense and form to meaningful and authentic English texts. For the first few months, their focus on form is combined with some tasks and activities from the set of course materials that Joy and Pete have purchased for general use and with the tasks and activities of their first project.

We see similar teacher initiation and control with Mark. In line with his beliefs, he introduces his learners to the course materials, which have recently been revised in view of the second phase innovations and are used for the first time. He soon combines working with the regular course materials with a focus on analysing and interpreting short stories. It is Mark's attempt to maintain his desired focus on literature in the upper forms of grammar school education.

Gradual release of teacher control

A second likeness is that all of the three respondents ultimately aim at releasing their initial control and initiative. They have learners work on their own, either individually or in groups, and are looking for ways to stimulate learner initiative. The degree and direction of how they release control corresponds with the extent of autonomy they believe their learners are capable of handling.

Mark is most closely in control of what his learners learn and how they should go about it. He is clear about what is expected of his learners in no uncertain terms. But

also Mark releases teacher control when he has his learners analyse a novel on their own in five or six consecutive lessons.

Joy is similarly explicit about the conditions in which the learners are expected to do the tasks and activities she wants them to do. Unlike Mark, though, she is much more learner-oriented and tends to intervene with attempts to have her learners reflect on what they are learning and how they go about it. In test preparation, she appears to go for a more dialogical and reflective approach than Mark does. Joy is generally focused on exploring what learners can do without her help. Once she sees and feels a group is doing a good job, she releases control and focuses on individuals or groups that need explicit encouragement or help. However, when groups are not getting any work done, she will surely tell them what they are expected to do.

Pete tends to always address his learners with attention and ease. Of all of the three respondents, Pete seems to most closely identify with his learners and the positions they are in. He does so with a sense of realism. He feels there is so much more to his learners' lives than just school. Pete seems successful in taking away pressure, if needed. He wishes to teach in a relaxed manner, and expects his learners to learn in similar ways. Nevertheless, also Pete can be very explicit about how learners are expected to behave. Releasing teacher control seems to come quite natural to him. What Mark, Joy and Pete all, to varying degrees, share is that the release of teacher control should be gradual. As teachers they all value opportunities to help or interfere whenever their assistance or mediation are required.

Full learner responsibility not a feasible concept

A third prominent resemblance, which logically relates to the conformity discussed above, is that the respondents feel that full learner responsibility is not a feasible construct in upper secondary education. The learners are generally felt to lack the knowledge, skills and mindsets to be held fully responsible for what they learn and how they go about it.

Joy claims that responsibility for her own learning only began to emerge when she was a university student. Joy feels responsible for exploring ways that help her learners to regulate their own learning. She does not expect her learners to assume this responsibility all by themselves. Even though Mark's grammar-school learners are generally gifted, he does not take learner responsibility for granted either. This is illustrated by the ways in which he monitors and controls how the learners carry out the tasks or activities they are asked to do or set themselves.

When learners are told to work individually in silence, Mark sees to it that they actually do work in silence. When learners work on their own in small groups, he has the groups think of aspects such as individual participation, planning and reflection on the activities in group logs, which have to be handed in after each session. Extrinsic incentives, such as the teacher determining and monitoring the conditions of a task or activity that has to be carried out, often play a role in Mark's didactics.

We have already mentioned that Pete most closely attempts to identify with his adolescent learners. He acknowledges that even though most of his learners know they are expected to get some school work done, many of them do not get down to giving it the efforts that are required. He assumes responsibility in the sense that of all of the three respondents, Pete most explicitly focuses on stimulating the learners' intrinsic motivation to get some school work done. He designs projects, tests, tasks and activities the learners can relate to and generally seem to like doing. We can add that Pete is always keen on taking away any pressure his learners might feel.

Learning by doing

A fourth and final resemblance is that the respondents feel that learner autonomy relates to tasks and activities the learners have to *do.* This means that learning how

to communicate in a foreign language invariably involves tasks and activities in which English has to be used receptively and productively.

Joy refers to the aspect of reflection on what learners have to or want to do. Mark refers to the competence to balance learning and doing activities. Finally, Pete refers to the learners' acceptance that English is a subject one learns by doing. This final resemblance in the ways in which the three respondents interpret learner autonomy directly relates to the construct of communicative competence in foreign language education, which we will discuss later.

Before relating our respondents' interpretations to some of the main theoretical issues regarding LA, there is a question that begs to be answered. In how far do we find converging evidence of the way in which a respondent interpreted LA at the beginning of the school year within and between the respondent's data gathered in the course of the school year?

11.2.3 Convergence of LA evidence

The data we gathered throughout the school year centred around three tests. The teachers had been asked to select and discuss three written tests they considered illustrative of their views of LA, CLE and/or effective testing. The interviews we had on each of the tests were again semi-structured and included open-ended questions that were meant to once more elicit the informants' views of the three constructs under discussion (see appendix xx). The evidence was later analysed for convergence of a particular construct interpretation expressed at the beginning of the school year.

We will present evidence as it converged or did not converge in the three tests in two columns. The column headed as *convergence* lists evidence that fully or partially illustrates the way in which a respondent interpreted a particular construct at the beginning of the school year. The column called *non-convergence* provides information on the absence or partial absence of a respondent's construct interpretation at the start of the year.

In this section we will focus on converging and/or non-converging evidence of Joy's Mark's and Pete's views of LA. Joy is the first informant we discuss. In the table below, DN stands for the *Dear Nobody* writing test, LB for the *Little Boy* grammar test, and finally IQ for the *Irish Question* reading comprehension test.

JOY	Convergence	Non-convergence
LA involves learner reflection on: -what they are doing; -what they have to do; -what they want to do; -what is required to achieve this.	Reflective open-ended questions on content (DN, IQ) or form (LB) always part of tasks preparing the learners to a test; Learners study basic self-assessment skills for writing, which allegedly they find difficult to apply; Learners repeatedly told to reflect on mistakes/errors they have made→ whether the learners follow up on her advice considered their responsibility.	No specific evidence of Joy's four "whats" of learner reflection in the follow-up data; Absence of reflective questions on form in the tasks preparatory to the DN-writing test; Joy experiencing difficulties in specifying what knowledge, skills or understandings are measured by a test; Test evaluation primarily done by Joy herself.
Role of the teacher:		
gradual release of teacher control; Full learner responsibility hard to achieve.	All of Joy's tests are characterised by firm teacher direction and control of potentially engaging tasks, followed by teacher-directed group assignments; Joy feels learners do not: - reflect on mistakes/errors made; - or cannot assume responsibility when they are given more autonomy in task performance (The Cal-test).	Joy does not release control in a linear way. If learners do not assume the kind of responsibility that Joy expects them to, she will revert again to more teacher direction and control.

Table 11.2 Convergence of Joy's interpretations of LA

Joy's interpretation of LA in terms of learner reflection at the beginning of the school year was challenging and ambitious. However, the data have shown that a focus on learner reflection is far from easy to realise. Joy does indeed focus on reflective questions, but they are not of the kind she mentioned before the school year started. The reflective questions asked in the preparatory tasks to the tests are related to communicative content ('What do you think of Chris's reaction to the news?' or 'What's the function of this paragraph?') or form ('Why do you think this particular tense and/or form must be used in this instance?'). Open-ended questions such as these, tend to refer less directly to overall learning than the learners' themselves reflecting on what they 'are doing, …have to do, …want to do' and 'what is required in order to achieve this'.

Even though Joy's interpretation of LA in terms of learner reflection did not converge in a convincing way, we would like to refer to her focus on reflection as genuine and sincere.

Joy seriously attempts to have her learners assess and evaluate their own writing. Besides, she persistently tells her learners to focus on mistakes and errors

they have made. Despite her efforts, Joy states a dilemma she cannot solve in any easy way: her learners do not or cannot reflect on their errors/mistakes and tend to be product- rather than process-oriented. In our final chapter, we will return to the dilemma stated by Joy.

Prerequisite to Joy's reflective interpretation of LA seems that she herself cannot, or only to a limited extent, come up with answers to the reflective questions she considers helpful to the learners and supportive to fostering their own autonomy. It appears to be far from easy for Joy to indicate what knowledge, skills or understandings are required to do well on a test.

Joy's release of teacher control does not seem to go any further than allowing her class to work in groups on a regular basis. In the *Irish Question* project, the learners were generally given more autonomy. Joy, however, regretted that she had given her class more liberties. It appeared that only a few learners were able to plan and study for the project tasks in a serious way. This seems to indicate that, from Joy's perspective, the assumption and release of teacher control is a balancing act.

We will now turn to Mark. He told us at the start of the school year that LA involved planning skills, self-assessment and evaluation skills and the ability to balance learning and doing activities. We will concentrate on the convergence and non-convergence of these aspects in Mark's assessment and evaluation practice in the course of the year. In Mark's tables DW stands for the *'A Day's Wait'* literary test, UF for the *Unicom Finals unit 8* test and finally PA for the *Practical Assignment* test, which required groups of learners to analyse a novel on their own.

Mark	Convergence	Non-convergence
LA involves three competences:		
planning skills;	Only the PA test requires a relatively high degree of planning skills, closely monitored by Mark by way of the group logs he had his groups write.	No explicit focus on planning skills in preparation for the DW and UF tests: the learners intended to be doing what the teacher tells them to do.
self-assessment and evaluation skills;	Three factors might implicitly have led to the development of self-assessment and evaluation skills: - Mark being relatively explicit about the knowledge, skills and understandings required to do well on a test; - the PA test requiring learners to monitor their own progress. - the assessment criteria explained and handed out before groups worked on the PA test.	No explicit focus on self-assessment and evaluation skills in preparing learners for the DW and UF tests; To Mark's regret, there was no time to discuss any of the tests in class after they had been returned.
ability to balance learning and doing.	The fact that Mark coherently distinguishes between learning and doing activities in his own didactics, might have led to learners doing the same.	None of the tests show that the balance between learning and doing activities was distinctly focused on.
Role of the teacher:		
Initially strong teacher control;	Full convergence: all of the data show that teacher control, and some coercion if need be, are typical of Mark's approach.	
Full learner responsibility a step beyond self-regulated learning.	Relatively high degree of self-regulation in the PA test.	Learner responsibility seemed to be taken for granted in the DW and UF tests: no direct focus on self-regulation.

Table 11.3 Convergence of Mark's interpretations of LA

In reading the table above, it seems inevitable to conclude that Mark's initial interpretation of LA in terms of three competences was only reflected to some extent in the data provided by the PA test. Little or no evidence was found in the data of the DW and UF tests.

The PA test was the only one to require a relatively high degree of planning skills. Besides, it seemed to hold the promise that having the groups self-monitor their own progress might lead to the development of self-assessment and evaluation skills. However, in none of the tests any evidence was found of a focus on learner awareness of the need to balance learning and doing activities, and of becoming a more autonomous learner by doing so. In addition, Mark claimed that time pressure had forced him to discard test evaluation and discussions of the tests in class after they had been returned. We will argue in our final chapter that formative as well as summative evaluation are key ingredients to fostering LA when learners learn to communicate in a foreign language.

So Mark's interpretation of LA at the beginning of the school year appears to have been rhetorical in part. This can be explained by referring to his strongly expressed disbelief that learners become more efficient and effective learners when they work and learn on their own. In the year of data collection, Mark's learners have basically acquired knowledge elements such as vocabulary, explicit grammar rules and literary notions by doing what Mark tells them to do: by memorisation and controlled practice.

On a more positive note, it is more than fair to add that Mark does indeed ensure that a lot of the ingredients to successfully learn English are there: high-level knowledge, which Mark seems to explain clearly and unambiguously to most of his learners, and the learners' application of that knowledge to related tasks. Mark's response to what exactly is required to do well on a test is generally elaborate and clear. This might have helped his grammar school learners to plan learning activities on their own and assess and evaluate their own progress.

In the next section, we will again return to the tests Joy had offered for discussion. We will now look at them from Pete's perspective on LA. In the table below, DN once more stands for the *Dear Nobody* writing test, LB for the *Little Boy* grammar test, and finally IQ for the *Irish Question* reading comprehension test.

Pete	Convergence	Non-convergence
LA involves didactics geared		
gradual increase of task and test complexity;	DN/LB: implicit in the tasks and activities preparing learners for the test;	DN/LB: no explicit evidence, mentioned by Pete; IQ: no evidence;
the learner's willingness to learn by doing;	DN/LB/IQ: - Pete offering advice and leaving it up to the learner to follow it up; - Learners know very well what they should do to learn effectively, but generally do not, e.g.: * work on their vocab by critically studying input; * review their grammar; * hand in their assignments; * discuss matters with Pete. * Learners must have the will and determination to practise and use what was demonstrated and practised in class.	
forms of peer learning.	DN/LB/IQ: implicit evidence: learners often work in groups.	DN/LB/IQ: no explicit reference to peer learning. mentioned by Pete
Role of the teacher		
gradual release of teacher control;	Pete active as a teacher, taking great care to: - motivate his learners from the start; - not to discourage his learners by tasks considered too challenging; - scaffold the level if need be; - teach learners how to reason why a particular verbal tense or form is correct in a given context; - demonstrate how he constructs reading comprehension questions as a teacher.	

Full learner responsibility a vague concept, beyond the reach of fourth-formers.	**DN/LB/IQ** **- Unrealistic to assume that adolescent learners have the responsibility to reflect on a test once it has been returned;** **- Pete responsible for what happens in the classroom, his learners for what occurs outside.**	

Table 11.4 Convergence of Pete's interpretations of LA

The learner's willingness to learn by doing and his/her responsibility to learn were two aspects of LA that frequently recurred in Pete's follow-up data. The two other aspects of LA Pete mentioned in the first interview, i.e. the gradual increase of task and test complexity and forms of peer learning, appeared to be aspects of his didactics that he did not explicitly mention in the follow-up interviews. Yet, they had clearly been part of Pete's didactics, as shown by our analyses of the tasks and activities preparatory to the tests and our limited classroom observations. Peer cooperation was prominent in the lessons we observed, albeit perhaps a little too self-evident. This might be an explanation why the increase of task and test complexity and instances of peer learning were not explicitly referred to by Pete in the follow-up interviews.

Studying the adolescent minds and behaviour of his learners, and, for that matter, the feelings and thoughts of his two sons who were working, and sometimes failed to work, their ways towards adult independence, has made him realise that for adolescents there is so much more to life than school. Besides, at school itself there are far more subjects than just English. It is a given fact to Pete that learners are quite capable of formulating what they *should* do to become more efficient learners. Yet, for a host of reasons they do not follow up on their teacher's advice or learn from what a teacher demonstrates to them or has them experience in class. Pete claims what is often missing is the learners' willingness to do what they are asked to do or to carry out what is likely to turn them into more effective and more autonomous learners. It is not entirely clear whether Pete accepts all this as a given fact or whether he considers this a dilemma. Whichever is the case, Pete feels his responsibility cannot go any further than the classroom. His learners are accountable for what happens outside.

Nevertheless, it is this very learner willingness on which his interpretation of LA seems to hinge. That seems to explain why Pete takes learner and learning motivation so seriously in the classroom tasks and activities he constructs. More so than Mark and Joy, Pete primarily aims at increasing his learners' intrinsic motivation. For fear of losing the intrinsic momentum, Pete is wary of tasks and activities that might be too challenging for his learners. Such a position might, at times, lead to over-simplification.

Yet, in our final chapter we will argue that there is and should be room for tasks that are challenging from a cognitive point of view as well. Joy, with her focus on learner reflection, and Mark, with his focus on knowledge-based skills, seem to be more aware of the cognitive and extrinsic aspects of learner motivation than Pete is. Pete, however, is unrivalled in his efforts to understand the adolescent mind and challenging his learners to learn for the fun of it.

The tentative conclusion we have just arrived at, paves the way for our argument that a teacher who manages to combine the qualities and perspectives of Joy, Mark and Pete, and perhaps adds some personal perspectives of his/her own, is likely to

become a teacher who is more capable of teaching learners how to learn to communicate in another language.

In the next section, we will relate the similarities as well as the different perspectives of our three informants to the contents of our theoretical chapter on LA. We will in particular go into what is there and what is missing and refer to the dilemmas a particular perspective might lead to.

11.2.4 The construct of LA and the LA data: what's there and what's missing

In chapter 3 we explored the construct of LA by concentrating on defining parameters of autonomy, discussing four arguments that justify a focus on learner autonomy in formal educational settings, and explored the relationship between learner autonomy and motivation. In this section we will primarily relate the LA data to the definitions, characteristics and aspects of autonomy we highlighted in chapter 3. This will be done under two headings: *what's there* and *what's missing*. Our theoretical discussion of autonomy helped to define what we actually found in the teacher data and what was missing. This section will provide input for our discussion of LA in the final chapter of this study.

What's there?

The respondents' concern with their learners' engagement and personal motivation to learn;

Joy, Mark and Pete are all keen on motivating their learners. Joy and Pete do so by constructing meaningful project tasks and activities, in which a host of mostly unspecified knowledge and skills are integrated. They are also keen on scaffolding English grammar, by reducing it to a primary focus on verbal tense and form and word order. Mark attempts to engage his learners in a completely different way. More so than Joy and Pete, Mark believes in building a firm knowledge base of discrete grammar points and memorised idioms that allows learners to interpret and appreciate literary texts. With a high-level focus on form and challenging literary texts, Mark hopes to positively affect his learners' engagement and personal motivation to learn English.

Teacher-initiated learning environments that encourage and enable learners to express who they are and what they think;

All of the three informants ultimately, and in many ways impressively, aim at helping and challenging learners to express who they are and what they think. They firmly believe in this ambitious objective, despite the different ways in which Mark on the one hand and Pete and Joy on the other attempt to realise this important target.

Teacher interventions if required;

Joy, Mark and Pete all believe in teacher direction and control, which is to be released only gradually. They intervene out of a responsibility both for the level of English and for the well-being of their adolescent learners. It helped both their learners and themselves survive in a turbulent year of curricular and didactic change.

Autonomous behaviour can neither be described easily, nor objectively;

It appeared to be far from easy for the respondents to specify learner autonomy or self-regulated learning. In defining LA, Joy, Mark and Pete focused on ways they felt would increase the autonomy of their learners. In their efforts to do so, they came up with interesting perspectives that all deserve to be followed up and investigated further. Yet, the perspectives were not developed in a structured way by the three respondents. Some of the perspectives even remained implicit.

Acknowledgement that LA is not a steady state;

LA is seen by the respondents in terms of task performance. As tasks differ, so may the performance of their learners. The degree of autonomy may also depend on the particular subject a learner attends. One excellent learner of English informally reported on her being notoriously bad at maths. A peer of hers who she helped out with English showed her similar courtesy when it concerned maths.

Evidence of a small number of, basically atheneum and gymnasium, learners who allegedly are autonomous learners, insightful and good at English for no apparent reason;

All of the three respondents reported on a minority of learners they felt to be more autonomous than others. These learners appeared to be motivated for English, good at it and more autonomous in task performance than others. None of the respondents could think of reasons why this might be the case. It nevertheless seems to corroborate the idea that there always have been autonomous learners, in any given educational context and that autonomy has been an educational objective in many contexts in the past.

The majority of learners allegedly benefiting from foundational teacher initiative and control;

It was frequently reported on by all of the three informants that teacher initiative and control is essential for adolescent learners. There were no reports of learners who all of a sudden started to flower and bloom when given the opportunity to work and learn on their own. Joy regretted having given her 4-atheneum learners too much freedom in preparing for the IQ and *Cal* writing tests, which she was going to remedy in time to come. Pete almost takes for granted that for adolescents there is so much more to life than just school. Mark's success with his PA test might, at least in part, be attributable to the close way in which he had the groups monitor their own progress and individual contributions. All of the respondents sympathised with their learners, who seemed to be overburdened by the school work they had to do. The respondents were always willing to help out any of their learners if need be.

Given the learners' age, the developmental stages they are in and their daily concerns and interests, a lot of learners allegedly lack the concentration, perseverance and stamina to regulate their own learning;

It is a belief frequently expressed by all of the three respondents. Pete, however, is most explicit on the matter.

LA is action-based, transfer-oriented and puts great demands on the knowledge and skills needed to carry out a task successfully as well as on the type of tasks, actions and activities to be carried out.

All of the respondents share a belief of learning by doing, and have transfer in mind when they have their learners perform tasks and activities. However, we see that Joy and Pete's efforts are more closely linked to meaningful communication than Mark's efforts are. Having said this, Mark's strong point is his focus on knowledge and his belief that knowledge is at the basis of all skills. Of all of the three informants, Mark is best at specifying what learners should know and be able to do to perform well on a test item or task. Joy and Pete clearly experience more difficulty in specifying what is required to do well on a task or test, particularly when it concerns the integration of knowledge and skills.

Attention to LA in the Netherlands does not fundamentally affect the informants' existing practical methodologies and techniques;

We have already reported on the dominance and recurrence of the beliefs that Joy, Mark and Pete had at the beginning of the school year. What they all attempted

to do was to fit in their existing beliefs and didactics as well as possible with the requirements of the renewed second phase. They attempted to keep up what they believed in, and what had worked for them and for their learners over years of experience.

What's missing?

The three academically-trained teachers referring to any theory related to LA.

None of the informants referred to academic theory in any specific or informed way. This relates to all of the theories and background information we dealt with in our chapter on LA. A scientific or scholarly background of the second phase curricular and didactic innovations, if at all present, did not filter down to the three academics attempting to teach 16-year-olds how to communicate in English and monitor their progress.

A majority of learners accepting responsibility for their own learning without teacher-initiation and control;

In the three contexts we studied, none of the respondents reported on learners spontaneously and self-evidently accepting responsibility for their own learning. This could be interpreted as self-fulfilling prophesy. After all, none of the informants believed in full learner responsibility from the start. In our final chapter we will put forward that such a view does not do justice to the serious problems our respondents signal.

Data on learner efforts to understand and explicate what one is learning, why one is learning, how one is learning and with what degree of success;

Both Joy and Mark promised to go for learner reflection and the development of language awareness at the beginning of the year. Yet, the data show that these perspectives were either partly missing in their assessment and evaluation practices, or that they proved to be too complicated or time-consuming to follow up any further. For Pete, any answers to the questions above were 'vague' from the start. Evaluation, as we defined it in chapter 5, was not a structural component of the respondents' teaching and testing practices.

Structural efforts of the respondents having their learners determine the learning objectives, define content and progression, select from useful strategies or techniques or share in any decisions and initiatives taken;

No specific evidence at all was found of the three informants deliberately attempting to have the learners share in any of the decisions.

Structural and detailed efforts to develop the learners' metacognitive and metalinguistic capacities (their ability to reflect on the learning process, the forms of the target language, and the uses to which the target language can be put);

In our next chapter we will argue that developing a learner's metacognitive and metalinguistic capacities ought to be a key focus both in researching and in developing learner autonomy in formal educational settings. Evidence of such a focus was weak and not always persistent in the cases of Joy and Mark. Pete does not appear to go any further than teaching learners how to reason why a particular verbal tense or form is the most appropriate in a given communicative context. Impressive, though, the effort is, we would like to argue it is not nearly enough if one aims to have young learners develop awareness of form, usage and use in English.

Structural efforts to have their learners plan, monitor and evaluate their learning both formatively and summatively.

The learners' responsibility does not seem to go any further than carrying out what has been planned on sheets handed out at the beginning of each term. As

already reported on, structural formative and summative evaluation of practice tasks and tests was, in general, missing in the data of this study. Joy was a consistent exception to the rule. Yet, her evaluations seemed basically teacher-directed and tended to favour error correction without the learners giving their mistakes and errors much thought, let alone reflecting on them.

11.3 Communicative language education

There is no learner autonomy without learning content of some kind. Over the years, second and foreign language education has increasingly been influenced by course materials and curricula that are defined as communicative.

At the beginning of the school year, our three respondents were asked how they interpreted communicative language education. Given the attention to CLE in the Netherlands since the 1980s, the question caused more difficulties than we had anticipated. Only Pete's response came with relative ease. After some hesitation, Joy essentially focused on the dilemma of balancing meaning and form. Mark set off by quoting John F. Kennedy and stressed that the ability to communicate is a life skill. Despite these hesitant starts, the respondents' interpretations of CLE appear to relate to important aspects of the construct as they have been mentioned in chapter 4.

Similar to our discussion of LA, we will first present a table with the respondents' initial interpretations of the construct. Then, we will have a brief section on two similarities of the ways in which the respondents attempted to define CLE. Next, we will turn to evidence of the respondents' CLE interpretations as it converged in the year of data collection. Finally, we will relate the data to essential parts of chapter 4 and again discuss what is there and what is missing.

11.3.1 The respondents' interpretations of CLE

We will start with an overview of what Joy, Mark and Pete read into CLE at the beginning of the school year.

JOY	MARK	PETE
CLE involves balancing meaning and form, characterised by:	**CLE involves direct instruction in form and usage, characterised by:**	**CLE involves the communication of meaning, characterised by:**
- effort to balance meaning and form; **-neither drills, nor exclusive focus on discrete grammar points;** **-agreement on what is communicative and what it is not.**	**-teacher instruction, engagement and admonishment;** **-a cooperative class responsive to the teacher's direct instructions;** **-learner awareness of contrastive form and usage.**	**-willingness to put messages across to an interlocutor who is willing to understand;** **-the messages put across being meaningful to the interlocutors.**

Table 11.5 The respondents' interpretations of communicative language education

Joy experienced most difficulty when she was asked to explain how she would define CLE. She basically focused on the dilemma of how to combine her primary focus on meaning and meaningful communication with a necessary focus on form, without resorting to an overemphasis on discrete grammar points. She was also bothered by the fact that there is so little agreement between colleagues about what is communicative and what is not.

After indicating that the ability to communicate is a life skill, a view that seems hard to find fault with, Mark stressed the directive and exemplary role of the teacher in his interpretation of CLE. The teacher speaks English, masters the language, and is aware of the problems that Dutch speakers experience when using English. Gradually, by being regularly confronted with teacher-initiated examples and being challenged or even a little forced to speak and write English, the learners develop awareness of contrastive form and usage. The basic requirement is to have learners who are responsive to the teacher's direct instructions.

Pete's response came with ease and deliberation. He was remarkably quick at defining a construct he had initially referred to as 'vague'. Pete stressed a person's willingness to put messages across in another language and the other person's willingness to attentively listen to and adequately respond to that message. He stressed the fact that it is important for messages to be meaningful to the interlocutors.

Again, the three different perspectives the informants adopt lead to interesting reflections on didactic realisations of the construct of communicative competence in formal foreign language education. Joy raises the dilemma of striking an effective balance between a focus on meaning and a focus on form and mediate what is communicative and what is not on the way. Mark stresses the important role a teacher has when he wishes learners to read, listen to, speak and write in English. He is convinced his learners need a teacher in an instructional and exemplary role. Mark also hinted at a class's willingness to cooperate with the teacher. Pete primarily concentrates on this very willingness in his interpretation. Rather than speaking of a whole class, he focuses on the individuals involved in communicative acts.

11.3.2 Similarities in the respondents' interpretations of CLE

Although there is, on the whole, less consensus in the informants' interpretations of CLE than was the case with their understanding of LA, we would like to highlight two similarities between the three informants. First, there is the difficulty in specifying what CLE involves. Second, meaning, form and usage are generally part of the respondents' interpretations of CLE.

CLE and communicative competence are constructions that are hard to define and specify.

Joy and Mark initially experienced difficulties in formulating what CLE involved. Despite the relative ease with which he offers his interpretation, Pete had frequently referred to communicative skills as 'vague' skills that are hard to specify. In our discussion chapter we will take up Joy's plea and attempt to determine in more detail what can be called communicative and what cannot.

All of the respondents, to varying degrees, refer to form, usage and/or meaning in their interpretations of CLE.

Joy wonders in how far form errors and knowledge of form may hinder or help effective communication. Mark pleads for the development of language awareness of English form and usage. Pete focuses on the meaningfulness of messages.

11.3.3 Convergence of CLE evidence

Below, we will present and discuss in how far evidence converged of the informants' interpretations of CLE.

Joy	Convergence	Non-convergence
CLE involves balancing meaning and form		
effort to balance meaning and form;	**DN: half-way focus on meaning ('content') and form ('technicalities')** **LB: reconstruction of verbal tense and form in a meaningful context.**	**IQ: rather haphazard balance of meaning and form items.**
neither drills, nor exclusive focus on discrete grammar points;	**DN,LB,IQ: fully convergent.**	
agreement on what is communicative and what it is not.	**DN: (sample) comparison of her own assessments with Pete's; both informants primarily assess holistically.**	

Table 11.6 Convergence of Joy's interpretations of CLE

The analyses and discussions of Joy's tests in the course of the school year amply illustrate what she reads into CLE. Her effort to balance meaning and form is both an objective and a dilemma. This is illustrated by the DN writing test and in the LB grammar test. The DN test impressively targets the learners' feelings and thoughts regarding teenage pregnancies. The learners' efforts are scored for meaning as well as for form. Joy appears to be scoring the test holistically rather than analytically. Her assessment criteria are only defined in broad terms. If in doubt about the fairness of her assessments, she consults with Pete on how he would score a particular part of a test.

Joy seemed a little embarrassed by the difficulties she had when defining what knowledge, skills or understandings were required to successfully do the DN test. It was one of the reasons why she chose for a discussion of the LB grammar test. The LB test had been administered some time before the DN test. She felt it was easier to specify what the learners should know and be able to do in the grammar test. The LB test illustrates how important it is for Joy to assess what she considers to be essential grammar in a meaningful (con)text. In the IQ test, the balance between meaning and form questions was more complicated and seemed rather haphazard. In assessing the construct of reading comprehension, it would seem more than interesting to reflect on the ways in which form aspects, such as essential grammar and vocabulary, affect the ways in which a learner interprets the meaning of a text.

The tests and the follow-up analyses and discussions convincingly show that Joy is not into drills and dealing with a host of discrete grammar points.

The next overview brings us to the evidence converging of how Mark takes to mean CLE.

Mark	Convergence	Non-convergence
CLE involves direct instruction in form and usage		
teacher instruction, engagement and admonishment;	**DW: fully convergent; English spoken a lot when dealing with literary texts;** **UF: fully convergent; Yet, now little spoken at all, let alone in English;** **PA: Mark instructing the groups in detail, presenting assessment criteria at the start, and monitoring progress by having the learners produce weekly group logs.**	
a cooperative class responsive to the teacher's direct instructions;	**DW, UF, PA: Mark very alert on learner attention and concentration; keen on motivating his learners from the start, intrinsically as well as extrinsically.**	
learner awareness of contrastive form and usage.		**DW, UF, PA: no explicit evidence of such a focus; developing awareness seemed largely left to the learners.**

Table 11.7 Convergence of Mark's interpretation of CLE

Developing learner awareness of contrastive form and usage seems a useful and interesting focus, both for LA and CLE. Unfortunately, no evidence was found of such a focus in Mark's data. Again, we feel it is important to stress that this does not in any way imply that Mark did not have such a focus in his lessons. Besides, two of the tests that Mark had selected were literary tests. These tests were all in English and had no explicit grammatical focus. The UF test did have a lot of discrete grammatical foci. Yet, this test was on unit 8. The UF course materials were organized in such a way that the odd chapters introduced new grammar and vocabulary and that the even chapters primarily repeated what had been explained, practised and studied before. The lesson observations show that the learners primarily worked individually and in absolute silence on the exercises preparing them for the UF test we discussed. Therefore, it would seem that developing awareness, if any, was largely left to the learners.

Next, we will present how Pete's constructions of CLE were reflected in the data.

Pete	Convergence	Non-convergence
CLE involves willingness to communicate meaning		
willingness to put messages across to an interlocutor who is willing to understand;	**DN: fully convergent→** **- test taps into the learners' willingness to express personal views, feelings and thoughts;** **- Pete basically assesses as an interlocutor willing to learn what his learners think and feel;** **- to keep up his learner's and his own motivation, Pete rarely ever uses the same test twice;** **- a lot of English input and all project tasks in English;** **- dictionary use and paraphrase promoted as strategies;** **- identification questions part of the practice tasks;** **IQ: test items Pete likes best are 'intelligent' questions that tap into a learner's understanding and interpretation of the reading comprehension text.**	**LB: had a grammatical focus, and did not aim at putting messages across.**
messages put across must be meaningful to the interlocutors.	**DN: most illustrative** **LB/IQ: illustrative in the sense that grammar and reading comprehension are assessed in a meaningful (con)text.**	

Table 11.8 Convergence of Pete's interpretation of CLE

The DN test best corresponds with how Pete interprets communicative performance in CLE. We have already mentioned that Pete regretted the discussion of the IQ test, which he considered to be a bad test. He was more keen on discussing the *Cal*-test, which was another writing test. That test was related to the Irish Question project as well.

Pete's preference for discussing the *Cal*-test instead of the IQ test is illustrative of a crucial belief of his. As a teacher, he wishes learners to develop opinions of their own on themes that are or can be made close to their hearts. As an assessor, he wishes to be entertained, impressed or moved by his learner's writing. This gives Pete's concern with the willingness to communicate and understand meaningful information an authenticity of its own.

The LB and IQ tests are less convincing illustrations of how Pete sees CLE. Nevertheless, the tests can also be viewed as meaningful on meaningful (con)texts. Moreover, Pete sees the LB grammar test as preparatory and crucial to having the learners develop their writing skills. The IQ test, according to Pete, illustrates well what test items he considers as utterly meaningless or irrelevant.

The ways in which Joy, Mark and Pete interpret and illustrate their constructions of CLE is more revealing than one would perhaps expect them to be at a first glance. In the next section, we will relate the CLE data to the contents of our chapter on communicative competence in foreign language education and again discuss what is there and what is missing.

11.3.4 The construct of CLE and the CLE data: what's there and what's missing

Chapter 4 first dealt with a methodological history of second and foreign language education. Next, we discussed developments that led to defining and further specifying the construct of communicative competence. Then, the focus was on three interpretations of communicative competence and some of their pedagogical implications. Finally, we dealt with some recent research findings. Again, the contents of chapter 4 have shaped our report on what we found and on what we did not find in the teacher data regarding CLE.

What's there?

Shared view of language as communication;

Joy, Mark and Pete ultimately aim at having their learners develop the ability to communicate in English on issues that are more personal and demanding than basic communicative functions, such as asking or showing someone the way, telling the time, making complaints, and the like.

Pete seems most aware of language being a social tool that speakers and writers use to make meaning. More so than Pete, Joy is wary of the precise role of form when learners learn to communicate. She objects to learners being 'unguided missiles' in the sense that they lack the grammar and vocabulary to coherently and consistently state what they want, think or feel in English. Mark firmly believes in laying a sound and elaborate foundation of form. He is of opinion that grammatical and idiomatic knowledge bases enable learners to effectively communicate in English, with little or no disturbing mistakes and errors.

Shared implicit belief in communication as the expression and interpretation of meaning;

Communication as defined by Savignon as a continuous process of expression, interpretation *and* negotiation of meaning is only partially covered by the data of this study. Negotiation is the missing link, which we will go into when we discuss what is missing in the teacher data.

Although the actual words expression and interpretation are not used by the three informants, there is convincing evidence that they implicitly believe in the importance and value of expressing and interpreting meaning when learners learn to communicate in English.

Yet, we have not come across convincing evidence that communication is viewed as a continuous process. What is also missing is explicit evidence on the role of negotiation of meaning. Meaning is often clarified or even made in dialogue with one another, with language being used as an artifact and a social tool.

Details on how knowledge and skills related to form are transferred from the teacher to the learner;

Mark relies on the course materials for laying the grammatical and idiomatic foundation allegedly required to successfully communicate in English. Knowledge and skills are basically transferred by the learners absorbing the teacher's instructions and explanations, doing exercises and learning idioms by heart.

Rather than focusing on a wide array of grammar points, Joy and Pete have singled out two areas of English grammar they feel are essential for effective communication: verbal tense and word order. Knowledge of verbal tense and/or form is transferred with the help of a heuristic *when*-scheme and individual and peer practice in which the learners have to explain which particular tense or form is the most appropriate in a given context.

Conflicting views on the role and use of published course materials;

Whereas Mark welcomes the course materials as helpful and decisive in determining the selection and gradation of course content, Joy and Pete soon decide that the set of course materials they use do not come up to their expectations. Instead, Joy and Pete prefer to construct their own projects, tasks and activities. This approach better corresponds with their beliefs in meaningful communication and learning and learner motivation.

Impossible to indicate one single approach or methodology that is convincingly more effective than others in teaching learners how to communicate;

In chapter 4 we have reported that the one single and most effective method does not exist. On the basis of the data of this study it is impossible to tell whether Mark's focus on discrete knowledge and literary interpretation or Joy and Pete's focus on integrative knowledge and skills is more effective in teaching learners how to communicate in English.

What we will argue in the final chapter, though, is that a combination of the qualities of the particular 'methods' of our three informants is likely to lead to a more effective approach in teaching learners how to communicate in English.

Remnants of the grammar-translation, direct and audiolingual methods;

The remnants mentioned above are predominantly found in Mark's didactic approach. In the data on the UF test we have seen that translations of Dutch into English and English into Dutch are prominent. We acknowledge there is and perhaps ought to be room for translation when learners learn to use another language. However, the test section of the UF test that meant to tap into writing skills appeared to be little more than a translation exercise in disguise. We will discuss this aspect in more detail when we review this test from the perspective of the knowledge, skills and insights it attempts to measure.

Mark also regularly stresses the importance of teaching in English and the learners responding in English. This characteristic of the direct method seems foundational in Mark's approach in teaching learners how to communicate.

The UF exercises and test items also illustrate a belief in audiolingual overlearning, repetition and memorization.

By contrast, there are hardly any remnants of the grammar-translation, direct and audiolingual methods. Joy and Pete feel these methods are too mechanical. They are more interested in project education and task- and/or problem-based learning, which are more recent approaches in foreign language education.

The above should not be seen as a disqualification of Mark's approach. Interestingly, Joy and Pete themselves seem to speak less English than Mark does in everyday lessons. It could also be argued whether the limited time Joy and Pete spend on grammar and vocabulary will suffice to enable learners to monitor and evaluate their own learning from grammatical and lexical points of view.

'From surface structure to meaning' and 'from meaning to surface structure' perspectives;

Mark's interpretation of CLE seems to correspond best with a *from surface structure to meaning* construction of communicative competence. Joy and Pete's views are more in line with a *from meaning to surface structure* interpretation. Attention to *specification of context* was largely absent. This will be discussed in the section on what is missing.

In our final chapter, we will contend that a focus on communication requires structural attention to all of the three perspectives of communicative language competence.

Primary focus on grammatical competence rather than on other areas of linguistic competence, such as lexical, semantic and phonological competences

Linguistic competence predominantly equals grammatical competence in our respondents' perceptions. Other areas of linguistic competence, such as lexical, semantic and phonological competences, appear to be concentrated on less intensively.

Yet, Mark pays attention to lexis, Joy tells her learners about the importance of building one's own vocabulary, and Pete points at dictionary use. Despite all this, the informants' attention does not seem to go any further than exhortations such as telling their learners to learn idioms by heart, construct their own word files, and look up any words they do not understand.

Towards the end of the year, Joy expressed that Pete and herself were far from pleased with the quality and range of vocabulary that their learners had managed to build up so far. Immersing learners English project tasks and texts did not appear to be nearly enough for the learners to master a fair amount of productive and receptive vocabulary.

Attention to discourse competence

There is some evidence of a focus on discourse competence. The respondents mention attention to word order, the ways in which sentences and paragraphs relate to one another in logic and style, and refer to skills such as note taking and making summaries. Besides, the use and interpretation of a variety of texts is prominent. However, grammatical competence seems to take precedence over lexical, discourse, sociolinguistic or strategic competences.

Frequent use of literary texts

Literary texts, plays and film scripts were prominent in Joy's, Mark's and Pete's teaching and testing practice. The texts allegedly engaged the learners and challenged them to identification and self-expression. In our final chapter we will go into the relevance of the use of literary texts when learners learn to communicate in another language.

Next, we will deal with what was missing in the teacher data on CLE in relation with some of the main issues presented in chapter 4.

What's missing?

Knowledge of academic theory and research outcomes on aspects and interpretations of communicative competence in second and foreign language education;

Again, the respondents never referred to any of the aspects and developments regarding CLE we discussed when other languages are taught and learned. It suggests an undesirable rift between the world of academe and everyday school practice.

Lack of specification of context and the unambiguous and clear statement of: communicative objectives, required knowledge and skills, functions and notions, and communicative teaching and assessment strategies;

The three respondents do not seem to think in terms of specific communicative objectives and targets set for their learners and/or themselves. Neither did the informants provide any detailed information on how *competence* elements relate to the *performance* of the tasks they wish their learners to accomplish. Functions, notions and strategies were only occasionally referred to and not paid attention to in a structural way.

An explicit focus on the fact that communication requires negotiation;

We have already referred to Savignon's definition of communication above and claimed that the aspect of negotiation was largely missing in the teacher data. Sometimes, meaning in communication can be crystal clear. This is particularly so when two or more people share the same sociocultural backgrounds and the willingness to understand one another.

More often than not, though, meaning in verbal or non-verbal communication is less self-evident. We would like to argue that such cases require mediation and negotiation, and that in this process meaning or different shades of meaning are created.

Both meaning and mediation will be two important factors of a didactic approach we will suggest when teachers attempt to teach learners how to communicate in another language. We would like to argue that the process of mediation and negotiation does not only concern what is meant (meaning) but also how it is expressed (form). More of this in our final chapter.

Attention to lexical, semantic and phonological competences;

As explained earlier, grammatical competence was most prominent when the teachers dealt with aspects of linguistic competence. In the final chapter, we will plead for a more even distribution of competences with the complex construct of communicative competence.

Attention to sociolinguistic, pragmatic and strategic competences;

Perhaps with the exception of discourse competence, the sociolinguistic, pragmatic and strategic competences that are part of communicative competence seem to play no or merely a subordinate role in the respondents' teaching and testing practices. Pete was the only one to explicitly mention awareness of register and the social status of the interlocutor as a characteristic of insight. Yet, in the tests we discussed, no evidence was found of such understanding.

Opportunities for the learners to practise their English and receive constructive feedback both on the ways in which they practise and learn and on the results of their efforts;

Lesson observations had confirmed the respondents' views that when working in groups not all of the learners succeeded in doing the tasks in English and did not come down or were unable to offer one another constructive feedback on what they were learning with what results.

Deliberate attempts to develop the learners' awareness of meaning, form and usage when they listen to, read, speak or write in English.

Joy's reference to balancing meaning and form and Mark's statement that he is used to paying attention to contrastive form and usage provided the only evidence that any of the respondents focused on the learners' awareness of meaning, form and usage in their teaching and testing practices. In the final chapter we will argue for

a more prominent position of this type of learner awareness when learners learn to communicate in another language.

11.4 Foreign language assessment and evaluation

We have now arrived at a discussion of the third and final construct of this investigation. In the semi-structured free-attitude interviews we had at the beginning of the school year, we asked several questions that elicited the respondents' views and beliefs concerning assessment and evaluation. First we asked our informants to define what they considered to be an effective written English test. Next, there was the question what knowledge, skills and insights of English the informants considered essential for the learners to master. Finally, we probed into their views on test washback.

In this section, we will look for convergence and non-convergence of the respondents' interpretations of effective testing, the knowledge and skills this involves, and the ways in which tests possibly affect how teachers teach and learners learn.

The structure of this section will differ somewhat from our previous sections on learner autonomy and communicative language education. This is because we raised three issues related to assessment, instead of merely eliciting interpretations on a single construct, as was the case with LA and CLE.

First, we will concentrate on Joy's, Mark's and Pete's interpretations of effective language testing and additionally discuss similarities and converging evidence. Second, we will concentrate on the knowledge, skills and insights the informants claimed that were important to be tested and the knowledge, skills and insight that they claimed to have tested in the three sample tests they had selected. Third, there is a section on the respondents' interpretations of washback, with its ensuing discussion of similarities and converging evidence. We will end our section on foreign language assessment and evaluation with a discussion of what is there and what is missing in the teacher data in comparison with some of the main issues raised in chapter 5.

11.4.1 Interpretations, similarities and evidence of effective language testing

The table below presents what Joy, Mark and Pete considered to be effective language testing at the beginning of the school year.

JOY	MARK	PETE
Effective testing characterised by:	**Effective testing characterised by:**	**Effective testing characterised by:**
-Level similar to or slightly higher than what a learner can be expected to handle; -Transfer: knowledge and skills applied to meaningful and authentic test tasks.	**-testing at the level the learners have attained; -test tasks that invite learners to use the grammar points and idioms practised, studied and learned in sentences of their own.**	**tests that are : - communicative; - meaningful; - in line with what has been taught and learned; - geared at transfer.**

Table 11.9 The respondents' interpretations of effective testing

Similarities

There are two similarities we wish to highlight before discussing any converging evidence. First, the respondents all include conceptions of validity in their

constructions of effective language testing. Second, all respondents refer to transference in their interpretations of an effective English written test.

The informants refer to the essential measurement quality of validity in indirect ways. Yet, we would like to argue, the references are there. Joy mentions the level a learner can be expected to handle as a quality of an effective test. By doing so, she indirectly refers to the level of the knowledge, skills and understandings that were typical of the tasks and activities that preceded the test. She wishes the level of the preparatory task and activities to relate to the summative test. Joy feels the learners should not be asked to perform in ways that cannot yet be expected of them. A test should be representative in that respect.

Mark also mentions the level the learners have attained as a quality of an effective test. However, Mark did not add Joy's reference that a test can also be at a slightly higher level than the learners are expected to handle. Joy literally referred to the phenomenon as the 'N+1 aspect', which is commonly associated with Krashen.

Pete feels it is important for a test to be in line with what has been taught and learned. It is both a more direct and a more general description of validity, as it can refer to many types of this essential measurement quality.

References to the transference of what has been learned is another common quality all of the three respondents identify.

Joy thinks it is essential for the knowledge and skills to be transferred to meaningful and authentic test tasks. Mark sees the transference of discrete grammar points in having the learners construct meaningful sentences of their own in which the particular grammar points are used correctly. Pete typically mentions that a test has to 'transfer something'. The word 'something' is characteristic here. We have already learned of Pete's references to complex constructs or integrative knowledge and skills as 'vague'. Nevertheless, Pete shares with his fellow informants a belief in the importance of transference in effective testing.

We will now briefly discuss in how far evidence was found of the respondents' constructions of an effective language test.

Converging and non-converging evidence

Joy

Evidence fully converged in the sense that Joy's tests amply illustrate her focus on transfer to meaningful tasks that are connected with authentic texts. This is particularly true for the DN and LB tests. Joy's and Pete's dissatisfaction with the IQ test already indicates that not all of the test items were considered meaningful. However, the Irish Question project, on which the IQ test was based was more than meaningful. It was characterised by the same variety in tasks and activities as the DN project on teenage pregnancies.

Joy's belief that the level of a test should be similar to the level of the preparatory tasks is also illustrated well. She takes great care in preparing her learners well for the DN and LB tests. The correspondence between test preparation and the actual test itself is utterly convincing. There is also an interesting correspondence of a validity issue that came up when we discussed the IQ test, which on the whole we would not label as a typically valid test. Joy chose the test, because it illustrated that learners who had prepared well and had done all of the preceding tasks and activities also did well on the test. The test results validated this aspect in Joy's perception.

What we did not find any evidence of was Joy's belief that a test can also be at a slightly higher level than what can typically be expected of them. The N + 1 puzzle remains unsolved.

Mark

Mark shares with Joy the careful way in which he structures the learning of his fourth-formers and creates the conditions that allow the learners to prepare for the

tests well. This is particularly true for the DW and PA literature tests. These two tests are illustrations of both texts and tests that the learners are likely to be able to handle. According to Mark, the learners have been dealing with four or five comparable short stories in class. The learners, therefore, had the opportunity to transfer the literary notions they had studied to a number of practice short stories.

However, regular school work primarily concerned the learners working their way through the solid, but fairly traditional UF course materials. The level of all of the UF tests administered in the course of the year seemed basically determined by the grammar, idioms and reading comprehension texts that were part of test preparation. What struck us was that the scores on the regular UF tests tended to fluctuate for quite a few learners. This was particularly the case with the marks the learners got for the reading comprehension tests that the UF tests usually ended with.

In the UF test we discussed, no evidence was found of the learners having to use grammar point and idioms in sentences of their own. The closest one gets to constructing sentences are the regular translation sentences from Dutch into English. This, however, is not what Mark meant by creativity at the beginning of the year.

Pete

In his qualification of what he felt to be an effective test, Pete had combined his belief in communication and meaningfulness with the essential test characteristics of validity and transfer. As a point of attention, we have already shown that Pete does not always know exactly what must be transferred.

The DN and LB tests convincingly illustrate Pete's beliefs that effective tests have to be communicative, meaningful, in line what has been taught and learned, and geared at transfer. The IQ test does not live up to Pete's standard of effective testing. That is why he feels the need to make some changes.

We will now turn to a discussion of the knowledge, skills and insights that the respondents feel are important to test and see in how far evidence emerged in the tests they had selected for discussion.

11.4.2 Converging evidence of essential knowledge, skills and inisghts

This what the informants mentioned about particular knowledge, skills and insights they felt were important for their learners to master and to be practised and tested.

JOY	MARK	PETE
Essential knowledge, skills and understandings:	**Essential knowledge, skills and understandings:**	**Essential knowledge, skills and understandings:**
Knowledge: **-irregular verbs and correct word order;** **Skills:** **-putting verbs into appropriate tenses and/or forms;** **-formulating questions and negations;** **Insight:** **-observable positive effects, but hard to specify and measure.**	**Knowledge:** **-knowledge and skills determined by the course materials;** **-knowledge of structural relationships between the words and phrases in an English sentence;** **Skills:** **-spelling ability;** **-ability to formulate precisely and appropriately;** **Insight:** ***-literary*: justifying personal opinions;** ***-grammatical*: understanding grammatical and/or semantic relationships.**	**Knowledge:** **-subject preceding main verb in English;** **-what is required to say and write what one wants to;** **-whether an English person would understand the utterances produced;** **Skills:** **-appropriate use of vocabulary;** **Insight:** **-grammatical insight;** **-some insight into register and the social status of the interlocutor.**

Table 11.10 The respondents' interpretations of key knowledge, skills and insights to be tested

Similarities

There are two similarities we would like to point at in the table on the knowledge, skills and insights. First, it appeared to be far from easy for the respondents to distinguish between knowledge, skills and understandings. Second, what was mentioned by the respondents largely related to aspects of linguistic competence.

The teachers had more difficulty in distinguishing between knowledge, skills and insights than table 11.10 suggests. The respondents first of all did not distinguish very clearly between declarative knowledge and procedural knowledge. Furthermore, the borderline between procedural knowledge, e.g. knowing how to spell, and skills, e.g. spelling ability, was often hard to demarcate when analysing the interviews. In addition, insight and understandings appeared to be hard to specify for all of the respondents. We have already mentioned and illustrated that Mark experiences fewer difficulties in mentioning what is measured in a particular test than Joy and Pete. One of the reasons might be that he is more oriented on discrete knowledge than Joy and Pete are.

In chapter 12, we will discuss the problems our respondents experienced in the light of the nationwide objectives, which are often formulated in terms of knowledge, skills and insight, and in the perspective of the popularity of competence- and/or performance-oriented assessment. The question we will attempt to answer is how important knowledge is in societies that so often refer to themselves as knowledge-based?

The second similarity between the three informants is that the knowledge, skills and insights that are mentioned largely concern aspects of linguistic competence.

The only exception to this are some of Mark's and Pete's statements. Mark also refers to knowledge of structural relationships between sentences and paragraphs, the ability to formulate appropriately, and literary insight. Pete intended to start with an enumeration of discrete grammar points, which he then generalised to 'what is required to say and write what one wants to', without further specification. It led to the claim that a learner should know whether an Englishman would understand what a learner has said or written. Pete interpreted insight as awareness of social status and the interlocutor, which he did not specify in any detail.

Converging and non-converging evidence

We will briefly overview in how far evidence converged of the knowledge, skills and insights the respondents considered essential for the learners to master.

Joy

DN: Partial convergence of the form aspects she mentioned. Form determined half of the final score on the writing test. There were no exact data on how the essentials mentioned by Joy affected her assessment.

LB: Full convergence: putting verbs into appropriate tenses and forms;

IQ: Non-convergence: test on reading comprehension was in Dutch.

Mark

DW: Partial convergence of the knowledge and skills to be determined by the course materials, the ability to formulate precisely and appropriately, and justifying personal opinions on literary texts;

UF: Full convergence of the knowledge and skills to be determined by the course materials, spelling ability, and partial convergence of understanding of grammatical and/or semantic relationships;

PA: Partial convergence, similar to DW above.

Pete

DN: Partial convergence of what is required to say or write what one wants to, whether an English person would understand the utterances produced, appropriate use of vocabulary and grammatical insight. No exact data on how the essentials mentioned by Pete affected his assessment.

LB: Implicit convergence of what is required to say or write what one wants to, and of grammatical insight.

IQ: Non-convergence: test on reading comprehension was in Dutch.

11.4.3 Converging evidence of test washback

Below, we present an overview of how Joy, Mark and Pete interpreted washback at the beginning of the school year.

JOY	MARK	PETE
Interpretations of washback:	**Interpretations of washback:**	**Interpretations of washback:**
-Tests not the instruments to help learners learn to communicate in English; **-Motivating practice tasks more likely to positively affect learning;** **-Marks important to learners, but a fallacy to believe learners only learn for marks.**	**-School tests necessary and essential incentives that influence how learners learn;** **-Tests a means to evaluate teaching and the teacher;** **-Positive test results stimulate learner and learning motivation.**	**-Tests not the ultimate goal, and should be subordinate to the process of acquiring language skills;** **-Marks can encourage or discourage learners;** **-An unsatisfactory mark might lead to learner reflection on what went wrong.**

Table 11.10 The respondents' interpretations of key knowledge, skills and insights to be tested

Similarities

Before turning to two similarities in the respondents' interpretations of washback, we will briefly ponder on a fundamental difference between them. Our data on the convergence of evidence of perceived washback effects on LA and CLE already illustrated that Joy feels that tests are not the instruments to help learners learn to communicate. Pete claims he could do without any tests altogether, and objects to tests being made the ultimate goal. Mark holds a view different from Joy's and Pete's. He stresses that tests are essential and necessary incentives that affect how learners learn.

Converging evidence on LA and CLE has already shown that the difference in interpretation of washback between the respondents at the beginning of the school year, appeared to be less prominent in the course of the year. For Joy, and according to her also for Pete, their interpretations of learner results affected their decision to pay more attention to vocabulary in times to come.

We would like to highlight two similarities. All of the informants acknowledge that marks are important for learners and that motivation is a factor in doing the tasks preparatory to a test and the test results a learner aims for.

All of the three informants first of all believe that marks are important to learners. Joy, however, feels that it would go too far to assume that learners are only willing to work for marks. In her experience, learners are more than willing to work at practice tasks when they regard them as motivating. Mark mentions the effect of positive test results on a learner's motivation to learn. Pete, however, also refers to the fact that insufficient marks can easily discourage learners to a great extent.

Despite these differences, by acknowledging that marks are important to learners and that they may affect a learner's motivation to learn the respondents potentially acknowledge the existence of washback. Yet, the perceived washback effects we reported on for three tests on LA and CLE were negligibly limited. Therefore, the question we will attempt to answer in our discussion chapter will not be whether washback exists, but how potential washback effects can be exploited to the full when learners learn to communicate in English. Data that illustrate the absence of a phenomenon or construct may be just as useful as data showing its presence.

We will now present in more detail how the three informants perceived washback in relation to the three sample tests they had selected.

Converging and non-converging evidence

Joy

DN, LB, IQ: Full convergence of the notion that tests are not the instruments to affect how learners learn, but that the tasks preparatory to the test are.

In none of the follow-up interviews, Joy offered examples that marks are important to the learners. Instead, she came up with examples of learners working well at the practice tasks for the IQ test and correspondingly doing well on the actual test.

Joy was the only respondent to consistently have the learners correct their tests and discuss the correct response in class. Yet, Joy frequently voices her dilemma. Learners generally do not reflect on the errors and mistakes made in a test.

Mark

DW, UF,PA: Non-convergence of explicit evidence. Mark's interpretation of test washback very much remained implicit or hidden. There is evidence, though, that he feels tests are important, because learners clearly work towards tests at regular intervals.

There is also the suggestion that the PA test may have a positive effect on how the learners learn and perhaps come to appreciate literature.

One of the reasons why evidence did not converge is that Mark appeared to be the only respondent *not* to discuss or evaluate any of the tests at all with his learners. He claimed he simply lacked the time and had to move on to the tasks preparing the learners for yet another test.

Pete

DN, LB, IQ: there is full convergence in all what Pete says and does regarding his belief that tests are not the ultimate goal and should be subordinate to the goal of acquiring language skills. This is particularly shown in the unorthodox ways in which he addresses testing. He hands out texts for the learners to study, before they are given the test questions the next time, or leaves the classroom when learners are doing a test.

There is partial evidence that Pete is aware of the fact that marks may discourage learners by establishing a bottom score of 5 for one his tests. Pete voices a dilemma similar to what Joy has put forward: learners know what they should do, but simply do not carry out what they should be doing. This is also true for test evaluation.

Pete feels responsible for preparing the learners in such a way that they acquire the necessary knowledge and skills. The learners are responsible for either taking up or discarding his advice to learn from errors.

Unlike Joy, there is little evidence of Pete paying a lot of attention to classroom discussion and evaluation of the tests.

In the next and final section of this paragraph, we will relate the testing data to some of the main issues presented in the theoretical chapter on foreign language assessment and evaluation.

11.4.4 Assessment and evaluation and the testing data: what's there and what's missing

In chapter 5, we first defined how we view the notions of testing, assessment and evaluation in this study. Next, we discussed four trends in the relatively brief history of language testing, i.e. the *pre-scientific, psychometric-structuralist, integrative-sociolinguistic* and *critical-dynamic* trends. Thirdly, we went into essential measurement qualities and explored the notion of test usefulness. Finally, we

discussed the relevance of foreign language assessment and evaluation of LA and CLE in formal educational settings.

In two sections to come, we will review what is there and what is missing in the teacher data in relation to the contents of chapter 5.

What's there

The respondents having a summative view of language testing;

All of the respondents, as well as their learners allegedly, see tests as summative events. A test closes off a particular stage of the learning process. This perspective seriously conflicts with views that language learning essentially involves cyclical processes, frequent negotiation and mediation, and the motivation not to be deterred or taken aback by any of the communicative problems learners will inevitably and necessarily encounter.

Therefore, we will contend that a summative and product-oriented perspective on assessment needs to be seriously addressed at any level by anyone having a vested interest in education. Summative and product-oriented perspectives on assessment do not do justice to developing the learner's autonomy in how (s)he learns to communicate in another language. We will go one step beyond our initial contention and argue that a more formative and process-oriented perspective on teaching, learning and testing will benefit all learners and all teachers of any given subject in formal education.

The respondents test and assess, but generally do not evaluate;

Joy, Mark and Pete test at regular intervals. These tests are generally planned at the beginning of the school year, as laid down in compulsory PTAs (plans for assessment and summative evaluation). The respondents assess in the sense that they classify 'a learner's knowledge and skills at a certain point in time by some test or procedure' (5.1 of this study).

However, the informants do not at all evaluate in the sense as we defined it in 5.1 as 'a retrospective and prospective procedure in which the results of an assessment are interpreted.

The respondents' focus on achievement testing rather than on proficiency testing;

In chapter 6 we distinguished between achievement and proficiency tests. Achievement tests are related to the instructions and contents of some deliberate teaching and learning process. Proficiency tests do not directly relate a given curriculum or course of study. Proficiency tests typically focus on future situations of language use and aim to measure performance-related knowledge and skills.

It was found that the tests of this investigation were achievement tests. They directly related to a preceding course of instruction and practice. We would like to argue for the need to have more formative proficiency tests that measure language ability in general and are not directly related to the curriculum. Thus, useful external measures are provided of the alleged proficiency of the learners.

Remnants of the pre-scientific trend in language testing;

In Mark's tests we find characteristics that are normally associated with pre-scientific times. The UF test relies heavily on translation from Dutch into English and English into Dutch. Four out of the seven test tasks are translation tasks that rely heavily on grammatical and lexical knowledge. The oral skills of speaking and listening generally seemed to be ignored.

Joy's and Pete's assessment procedure of the DN test is, partly at least, typical of the predominantly holistic assessment of pre-scientific times, in which judgment is in the hands of an experienced teacher who can tell what mark to give afer a few minutes' conversation or after reading a learner's essay or letter.

Characteristics of the psychometric-structuralist trend;

Four characteristics of Mark's assessment practice fit in with the psychometric-structuralist trend in language testing. The UF test is an instance of a discrete-point test. In addition, his concern with objective measures is a characteristic that typically fits in with this trend. The UF test also included a short-item reading comprehension test, which is not uncommon to psychometric structuralism. Besides, Mark's UF test is primarily an indirect test, i.e. a test ' that attempts to measure the alleged abilities underlying a skills, rather than testing the performance of the skill itself.' (5.2.2 of this study).

Joy and Pete are also concerned with reliability issues. They are very much aware that they assess holistically. They attempt to make their assessments more objective, reliable and valid by analysing each other's response, with the aim to arrive at comparable and fair scores and grades. Joy and Pete are also keen on any comments their learners make on the ways in which they have been assessed and the fairness of their assessments.

Characteristics of the integrative-sociolinguistic trend;

All of the three informants test writing performance integratively, i.e. by way of tests in which knowledge of the various components of the language system have to be integrated with an ability to produce and interpret language appropriately in its linguistic context Integrative tests and integrative thinking are typical of the integrative-sociolinguistic trend. In addition, Joy and Pete display awareness of the relevance and variety of social context of both the texts they choose and the tests they construct.

Communicative writing tests;

Joy's and Pete's DN test and Mark's DW and PA tests can be seen as communicative tests in the sense that they are performance tests that engage the learners in extended acts of communication and in which the learners have to assume social roles they are likely to assume in real-world settings.

It is no coincidence that the respondents respectively like their DN, DW or PA-tests most of all. It is yet another illustration that the teachers share a view of language as communication, even though the ways in which they have their learners prepare for the test differ considerably.

Joy's and Pete's DN test appear to best correspond with the argument that communicative tests need to be both system- and performance-oriented. However, for detailed information on both the system and the performance, we would rather turn to Mark.

Basic concerns with reliability as an essential measurement quality;

We have already discussed the informants' concern with reliability as a characterictic of the psychometric-structuralist trend of language testing. Here, we would like to stress that all of the three informants interpret reliability in terms of scoring and scorer reliability. In the section on what is missing, we will go into reliability concerns the teachers did not appear to be paying attention to.

Basic concerns with validity as an essential measurement quality.

A basic definition of test validity is deceptively straightforward. A test is supposed to be valid if it assesses what it is expected to assess. The three informants are all committed to assessing what they want to assess. What the respondents test and how they do it is first of all overwhelmingly in line with their beliefs. How learners are prepared to the tests, what is tested, and how this is done is generally in line with the ways in which they interpret the central constructs of this study.

The tests that converged best with the respondents' fundamental beliefs and interpretations of, most notably, CLE and to a lesser extent of LA, were Joy's and Pete's DN and LB tests, and Pete's DW and PA literary tests. It must be said, however, that the respondents' construct interpretations appeared to be somewhat limited in view of the theory we explored in this study.

Yet, the logic of the definition of validity we started this section with conceals an essential measurement quality that is more complicated and challenging than is apparent at first sight. Validity is more and more discussed in terms of construct validity. Knowledge of and a specific focus on construct validity was missing in our teacher data.

In the final chapter, we will look into ways in which the teachers can widen as well as deepen their interpretations of LA, CLE and assessment and evaluation by enhancing the construct validity of their assessment and evaluation practices. Thus, what is missing in this study may be just as revealing as our findings of what is there.

What's missing

Knowledge of academic theory and research outcomes on foreign language assessment and evaluation in second and foreign language education;

Again, no references to academic theory were made.

Formative and summative evaluation;

We have defined evaluation as a retrospective and prospective procedure in which the results of an assessment are interpreted. The informants do not or only rarely evaluate, neither summatively, nor formatively.

Positive washback effects of the language tests;

The potential effects that tests have to positively affect teaching and learning are not exploited by the respondents.

Characteristics of the critical-dynamic trend;

Above we have reasoned that particular features of the prescientific, psychometric-structuralist and the integrative-sociolinguistic trends come to the fore in the respondents' assessment practices. However, little or no features were found that can be called typical of postmodern thought, apart from a serious concern with test fairness by Joy and Pete, and by the critical concern of the teachers with the national reading examinations.

The five concerns with which Spolsky (1999: 702) typified postmodern language testing were largely missing. The concerns were related to (1) language testing ethics, (2) specification of test purpose and design, (3) the need for procedures to be open for inspection and feedback, (4) a recognition of professional standards, and (5) a preference for multiple methods of assessment and evaluation.

In our discussion chapter we will highlight two concerns in particular: formative assessment and evaluation and more dynamic and more dialogical forms of assessment.

Extended reliability concerns;

The teachers focused primarily on scoring and scorer reliability. The data did not reveal detailed information on other reliability issues. In our final chapter, we will discuss how the respondents' concerns with the reliability of their assessment and evaluation practices can be improved any further.

Extended validity concerns;

The respondents were predominantly concerned with issues related to content validity. In our discussion chapter we will plead for attention to the notion of construct validity, which pertains to the ways in which test scores are interpreted.

Both the extended validity and reliability concerns will be discussed in relation to other aspects of test usefulness, such as authenticity, interactiveness, impact and practicality.

11.5 Summary

In this chapter, we have reduced, compared and contrasted the data of *Testing for Autonomy* and related our findings to the constructs of *learner autonomy, communicative language education*, and *assessment and evaluation* as they have been discussed in chapters 3, 4 and 5. We particularly highlighted the similarities between the respondents and united their perspectives, without losing track of differences in interpretation.

The teacher data has been discussed as follows. For LA and CLE, we started with the informants' construct interpretations elicited at the beginning of the school year. Then, we discussed similarities in the ways in which the teachers viewed a construct. Next, we considered in how far evidence converged in the tests and subsequent interviews in the course of the school year. Finally, we compared our findings with the construct theory we presented earlier in the study. We commented on the presence as well as the absence of essential parts of our theoretical chapters under the headings of *what's there* and *what's missing*.

Our discussion of the construct of foreign language assessment and evaluation was structured somewhat differently. Here we went into three aspects of assessment and evaluation the study elicited. We presented the respondents' views on what they felt to be the qualities of an *effective English written test,* the *knowledge, skills and insights* they considered essential for their learners to master, and their constructions of test *washback.* Each of these aspects was discussed in terms of similarities and convergence and non-convergence of evidence. Finally, the findings were again united and discussed in terms of *what's there* and *what's missing.*

CHAPTER 12: AUTONOMY TESTED: A DISCUSSION

12.1 Introduction

This chapter ends our quest for autonomy in settings where adolescent learners learn to communicate in another language. In this introduction to the final chapter, we will first look back on the contents of the previous chapters. Then we will return to the research questions formulated at the start of this study and establish which questions have already been answered and refer to the ones that remain to be addressed. Finally, we will present the structure of our final chapter.

What went before

The investigation started with a chapter in which we explained our interest in the three central constructs and looked ahead at the chapters to come.

In chapter 2, we provided details on the exploratory multiple-case study. We first went into its rationale, objectives, questions, and method. Next, information was provided on how the teacher respondents had been selected, the five stages in which the study was carried out, and the ways in which the investigation was documented, analysed and reported on. We then discussed the three central constructs of our investigation from a theoretical point of view in three consecutive chapters.

Chapter 3 focused on Learner Autonomy (LA) as a pedagogical construct. Here, we addressed three questions: what LA is, why it can be seen as a viable pedagogical construct and goal, and what the relationship is between learner motivation and LA.

In chapter 4 we discussed the construct of communicative competence in foreign language education and explored some of its backgrounds. We started with a concise methodological history of second and foreign language education. We stressed the importance of neither rejecting the past, nor uncritically accepting the future as far as methodologies for the teaching and learning of another language were concerned. We then explored methodologies that aim at meaningful communication and a specification of the knowledge and skills required to communicate correctly, efficiently and appropriately. Such approaches aim to engage learners in communication, defined as “a continuous process of the expression, interpretation, and negotiation of meaning” (Savignon, 1997:14). This was followed by a discussion of influential models of communicative competence and three interpretations of the construct, characterised by a focus on form, meaning or context. We ended our chapter with some of the more recent research insights in communicative ability.

Chapter 5 dealt with the construct of foreign language assessment and evaluation. We first stressed the importance of clearly distinguishing between the notions testing, assessment and evaluation. In this study we defined assessment as any measurement of knowledge and skills at a given time and evaluation as the retro- and prospective inferences made from the results of a particular assessment procedure. Secondly, we discussed four trends in the field of language testing, subsequently the *pre-scientific*, *psychometric-structuralist*, *integrative-sociolinguistic*, and *critical-dynamic* trends. In a third section we explored three subject domains that have helped to establish professional standards in the field of language testing. We first discussed tests in terms of test purpose and test types. Then we went into the essential measurement qualities of reliability and validity. Finally we dealt with Bachman & Palmer’s definition of test usefulness as the sum of six interrelated qualities, i.e. *reliability*, *construct validity*, *authenticity*, *interactiveness*, *impact* and *practicality*.

In chapter 6 we briefly interrupted the momentum of our study with information on the context of secondary and teacher education in the Netherlands amidst turbulent

curricular and didactic reform. We then reported on the teacher data in four consecutive chapters.

Chapter 7 was a predominantly narrative chapter, in which the three respondent teachers, their core beliefs and their construct interpretations were introduced. The chapter was based on the interview data gathered before the start of the school year.

Chapters 8, 9 and 10 reported on the data collected in the course of the school year, when three sample written English tests were selected by and discussed with the three informants in the light of the research questions. Chapter 8 focused on *Joy, the budding professional*, chapter 9 on *Mark, the literary master*, and chapter 10 on *Pete, the project man.*

In chapter 11, the data of this study on LA, communicative language education, and foreign language assessment and evaluation was reduced, compared and contrasted by way of cross-case analyses, and subsequently related to the contents of the three theoretical chapters.

The research questions

Chapters 3 up to and including 11 have already provided answers to most of the research questions we raised at the start of this investigation. We will briefly return to the questions to establish which of them need to be addressed in this final chapter. Initially, three research questions were formulated.

1. What are the beliefs of degree-one teachers of English with regard to *learner autonomy*, *communicative language education,* and the role of *assessment and evaluation*?

2. How are their beliefs reflected in their assessment and evaluation practices?

3. What recommendations can be made to enhance learning environments in which tests have beneficial washback effects on the ways in which learners learn to communicate in English?

The generally formulated questions came to be specified in seven related questions. As the first four questions below have already been answered in the course of this study, we will add references to the chapters in which a particular question was addressed. The questions that remain to be answered will provide the structure of our final chapter.

1. What can we learn from studying theory on *learner autonomy, communicative language education,* and *foreign language assessment and evaluation*? **(chapters 3, 4 & 5)**
2. How do teachers of English define and specify *learner autonomy, communicative language education,* and *effective foreign language testing*? **(chapters 7 & 11)**
3. What beliefs, experiences and arguments underlie the design and administration of three sample informal tests constructed by the teachers themselves? **(chapters 8, 9, 10 & 11)**
4. How do teachers interpret the test results of the sample tests in relation to their views, opinions and beliefs of communicative language education and learner autonomy? **(chapters 8, 9, 10 & 11)**
5. How do the research findings compare and contrast with theoretical insights in *learner autonomy, communicative language education,* and *foreign language assessment and evaluation*? **(chapters 11 & 12)**

6. In how far has the second-phase reform been conducive to fostering learner autonomy, enhancing communicative language ability, and promoting effective assessment and evaluation? **(chapters 6 & 12)**
7. What can be learned from the investigation in view of further research and educational programmes on how to create positive washback of assessment and evaluation practices on classroom teaching and learning? **(chapter 12)**

What is to come

In this concluding chapter, we will primarily concentrate on answering research questions 5, 6 and 7. Response to these questions will be integrated into the three main sections of this chapter.

We will start with a section called *Theory and the teacher data*. Here, we will address the question how the research findings compare and contrast with the theoretical insights discussed in the chapters on learner autonomy, communicative language education, and foreign language assessment and evaluation.

We will do so by discussing for each of the three constructs: (1) core theory that helps to clarify and/or further explore the construct, (2) links between the teacher data and the theory discussed, (3) gaps that exist between the theory and the educational practices under investigation, and (4) an analysis of why these gaps appeared as they did.

In the second section named *A context of innovation and reform*, we will take up the question in how far the second-phase reform has been conducive to fostering learner autonomy, enhancing communicative language ability, and promoting effective assessment and evaluation. Here, we will focus on the specific context of secondary education in the Netherlands at the time the study was carried out.

In the third section entitled *Recommendations for pre- and in-service teacher education* , we will address the question what can be learned from the investigation in view of further research and educational programmes on how to create positive washback of assessment and evaluation practices on classroom teaching and learning in settings where learners learn to communicate in another language.

We will finish our chapter with a discussion of the *shortcomings* of the investigation and the dilemmas that had to be resolved in the course of it. The insights gained may be helpful in view of future research in subject areas related to the three constructs studied in this investigation.

12.2 Theory and the teacher data

We will subsequently refer to core theory from chapters 3, 4 and 5 that has helped us interpret the teacher data on the three constructs of this study, discuss links between the respondents' beliefs and construct interpretations, go into gaps between the theory discussed and the teacher data, and finally arrive at an analysis why these gaps appeared as they did.

We will first deal with LA, then discuss CLE, and finally ponder over foreign language assessment and evaluation.

12.2.1 Learner Autonomy

Core theory

Four theoretical strands from chapter 3 appeared to have been particularly useful in interpreting our LA teacher data. They subsequently were the various ways in which LA had been defined, two philosophical accounts of autonomy that highlighted the importance of the reasons that make human beings act, four perspectives on

human learning, and finally theories and notions related to learner and learning motivation.

First, there was the discussion of key definitions of LA that resulted in what we called parameters of autonomy. Our argument was that if teachers concentrate on one or more of the parameters below and involve their learners in any choices that can be made, LA is likely to increase. The parameters we identified were the:

- ✓ *engagement and motivation* to foster LA of both the teacher and the learner**;**
- ✓ creation and maintenance of a *challenging learning environment;*
- ✓ specific *knowledge, abilities and willingness* required to learn effectively***;***
- ✓ *learning objectives of the curriculum;*
- ✓ *curricular content and pace;*
- ✓ *tasks, actions and activities;*
- ✓ *learning strategies and techniques* involved*;*
- ✓ level of *transference of learning tasks, actions and activities;*
- ✓ *acceptance of responsibility* by the learner*,*
- ✓ learner's *concentration, perseverance and effort;*
- ✓ opportunities for the learner for *self-expression by communication of feelings and thoughts;*
- ✓ ways in which progress is *monitored, assessed and evaluated;*
- ✓ roles of *metacognitive* and *metalinguistic reflection.*

A second theoretical strand that may help to promote LA are two of the four philosophical accounts of autonomy we highlighted. They are the so-called *reasons-responsive* and *responsiveness-to-reasoning* interpretations of autonomy. Both accounts seem to relate well to stimulating learners to reflect on the learning process and have them develop their metacognitive and metalinguistic capacities.

The *reasons-responsive* interpretation of autonomy claims that humans become more autonomous if the number of motives on which a person acts increases. It is suggested that learners who can think of a number of reasons why they should practise writing letters in English is likely to be more autonomous than a learner who simply do as they are told.

In a *responsiveness-to-reasoning* interpretation of autonomy, the sheer number of motives on which a person acts is less important. In this account, the quality of reasoning, even within one motive to act, determines the level of autonomy of a human being. Thus, learners whose main motive of writing English letters is learning how to express themselves and intelligibly structure their thoughts, may in this account be seen as more autonomous learners than the ones who can give a number of purely instrumental reasons why it is important to learn how to write in English.

We feel that if teachers stimulate learners to think of motives behind the language-learning related actions or activities they want them to do and pay attention to the quality of reasoning of a given motive to act, the learners' autonomy is likely to be fostered.

A third theoretical section relevant to our interpretation of teacher data on LA is the discussion of the humanistic, cognitivist , constructivist and sociocultural perspectives on human learning and the attention they typically pay to a learner's motivation to learn. Humanists focus on creating a safe and affirmative learning environment and positive teacher-learner interaction. Cognitivists opt for a more conceptual approach and highlight a person's innate cognitive drives and the motivation that comes from deciding for oneself what to think, feel or do. Humans define themselves by making their own decisions, rather than by simply reacting to others. In constructivist and sociocultural perspectives, individual choice and the nature of the social context are emphasised. Each and every person is motivated in

unique ways. Individual actions and activities are always embedded in a specific milieu and are therefore inseparable from their sociocultural context. We feel that knowledge of and experience with key theories on human learning are likely to sensitise teachers to the fact that there are multiple ways in which human learning and autonomy can be interpreted and stimulated.

A fourth and final strand of relevant theory is a follow-up section on learner and learning motivation, in which we discussed intrinsic vs. extrinsic (Deci & Ryan, 1985) and instrumental vs. integrative motivation (Gardner & Lambert, 1972). We also referred to attribution theory, which claims that learners who consider that success or failure is the result of their own efforts to take responsibility are more effective learners than those who attribute success or failure to causes external to themselves or beyond their abilities. Another theorist we considered relevant was DeCharms (1984), who claims that learners who control their own learning and originate their own behaviour are more successful learners than those who only act on stimuli external to themselves. In view of learner and learning motivation, we supported Van Lier's (1996) plea for more attention to factors of intrinsic (innate) motivation in secondary education, which seems primarily dominated by extrinsic (environmental) factors. We feel that a teacher's approach of learner and learning motivation will become more directional if they explore important notions of the construct of motivation.

In the next section, we will discuss potential links between the LA data and the four strands of theory outlined above. The rationale of this section is that we feel an important starting point for teacher change and professional development is to look for links between a teacher's core beliefs and interpretations of important constructs and core academic theory.

LA practice and LA theory: potential links

We would like to single out three links between the respondents' beliefs, construct interpretations of LA and the theory we discussed. We feel these links may potentially interest teachers to explore key theory related to LA.

First, all of the respondents see LA as a relevant pedagogical goal. They all teach whatever they feel is important with the aim to have their learners express themselves in English. In addition, the respondents all feel that given the ages and developmental stages the adolescent learners are in, full learner responsibility is not feasible. They believe teachers should initiate learning and learning behaviour and intervene if need be. This view fits in with two misconceptions mentioned by Little (1991) that in fostering LA the teacher is meant to become redundant and that any intervention on the part of the teacher may destroy whatever autonomy the learners have managed to attain.

Our data has shown how crucial teachers are in creating the conditions, learning environments and determining the subject knowledge and skills that learners need in order to develop their autonomy. The fact that the respondents see LA as a relevant goal may trigger their interest to sensibly and sensitively look for tasks and activities the learners can do on their own and to search for those for which they need assistance and support.

Second, all of the teachers proved to be genuinely concerned with the well-being and learning motivation of their adolescent learners. The code *motivation* was assigned most often when the teacher interviews were analysed. In the respondents' perceptions, this was an important aspect on which the success or failure of LA seemed to hinge. This particular finding links up with the attention we paid to learner and learning motivation in the theoretical chapter on LA.

Third, the teachers came up with interesting and challenging interpretations of LA when they were asked to define what it was. According to Joy, LA involved having the learners reflect on what they were doing, what they had to do, what they wanted to do, and on what was required to achieve all this. Mark defined LA in terms of three

knowledge-based competences, i.e. planning skills, self-assessment and evaluation skills, and the ability to balance learning and doing activities. Pete felt LA involved didactics characterised by a gradual increase of task and text complexity, stimulating the learners' willingness to learn by doing and engaging them in forms of peer learning.

In chapter 11 we have discussed that evidence of the respondents' construct interpretations did not always converge in convincing ways. This is a matter we will return to in our recommendations.

LA practice and LA theory: crucial gaps

In view of the theory we discussed, we feel three gaps between LA theory and the LA practices need to be addressed in particular. The first gap is that the respondents' construct interpretations of LA seemed to have been only marginally informed by academic theory.

Secondly, it appeared to be far from easy for the respondents to remain focused on and explore their perspectives on LA in their teaching and testing practices.

Thirdly, according to the respondents only very few learners appeared to accept responsibility for their own learning without teacher initiation and control.

In the next section, we will analyse and discuss the gaps we highlighted, with the aim to arrive at recommendations for pre- and in-service teacher education in 12.4.

Analysis

In this section we will analyse and discus why the three gaps between LA practice and LA theory appeared as they did.

LA theory and LA practices

The investigation has shown that our three respondents never explicitly referred to any academic theory related to learner autonomy. Joy, who had only just finished her degree-one teacher education course, nevertheless seemed to have been inspired by theory in her interpretation of LA from a metacognitive and motivational point of view. Yet, in the course of the year it appeared how difficult it was for her to follow up and further explore her interest in metacognition and learner motivation. In Mark's interpretation of LA his alleged focus on his learners' self-assessment and self-evaluation seemed promising. Yet, little evidence was found of such a focus, apart from the successful group logs he had his learners produce in preparing for one of the tests he provided. Pete felt peer learning was important, and did indeed often have his learners work in groups. Yet, *what* exactly was learned and *how* the learners were supposed to go about it remained 'vague', to use the term he frequently used.

We can therefore conclude that the respondents' academic knowledge of the construct of LA was neither extensive, nor very detailed. It is hard to foster autonomy if teachers have little idea of what autonomy entails and how it can be implemented in their everyday teaching and testing practices.

We can also conclude that even if a potential aspect or parameter is focused on by the respondents, it appears to be hard and time-consuming to explore and implement. In section 12.4 we will plead for more detailed attention to the fact that our respondents had limited knowledge of what the complex construct involved and found it hard to implement. The matter is particularly poignant, because we will arrive at similar conclusions when we discuss the constructs of communicative language education and foreign language assessment and evaluation.

We feel there are at least six reasons why theory on LA appeared to have played such a marginal role for our three respondents. First, the academic theory on LA that was focused on in the second-phase reform was neither broad nor deep. Second, from the theory that was focused on, only key notions appeared to have filtered down to our respondent teachers. Examples of such notions are self-regulated learning,

learner activation, learner responsibility or planning skills. Little was known about how these key notions would be interpreted by teachers and their learners in the first official year of second-phase reform in the Netherlands. Third, theory had provided little empirical evidence that a focus on LA was actually going to work, unless teacher roles and didactic approaches would fundamentally change. Fourth, when the innovations were first introduced, there was little time for teachers to reflect on central concepts, experiment with them and evaluate results. In many cases, the situation seriously aggravated when it appeared that teachers were actually given more and larger classes to teach when the reform was introduced. Fifth, second-phase innovations appeared to be based on the implicit assumption that teacher beliefs could be changed by introducing key concepts top-down. Research on teachers' beliefs suggests that it is very hard to influence the beliefs and related practices of either beginning or experienced teachers (Pajares, 1992, Carter & Doyle, 1995; Woods, 1996).

This study has shown that almost all of the respondents' core beliefs are reflected in their teaching and testing practices and that these beliefs relate to the ways in which they define the central constructs of this investigation. This finding is corroborated by literature on teacher beliefs and the impact of beliefs on teaching practices. It has shown that teachers' general conceptions directly shape the development of context-specific conceptions, which in their turn lead to the choice of specific teaching activities (Sercu & St.John, 2007:58).

We feel that teachers' beliefs must be taken very seriously, and should be taken as a starting-point for any process of innovation or change. This has not been the case in the curricular and didactic reform that was the backdrop of the present study.

Our sixth and final argument why academic theory received so little attention by the teachers is related to the respondents' willingness to change and develop as professionals. What about their motivation to reflect, change and develop their professionalism, as many fellow academics and skilled workers do? After all, if teachers are not willing to change, if they are not willing to assume their responsibility and autonomy, education is likely to become a dead-end street, whatever the contextual conditions are. We therefore feel that the teachers' own attitudes to educational change and professional development ought to be addressed as well.

In the next section, we will analyse and discuss a second gap between LA practice and LA theory. It was far from easy for our respondents to realise their specific interpretations of LA in their educational practices.

The respondents' perspectives on LA hard to realise

We admire our respondents for the interesting and challenging ways they interpreted LA, with Joy concentrating on metacognition, Mark focusing on knowledge-based competences and Pete directing himself to didactic procedures that would help his learners develop their autonomy. Yet, putting their ideas into practice was not unproblematic. We feel this was in part caused by the six reasons we offered for the gaps we found between LA theory and LA practice. However, this particular finding calls for more detailed analysis. We would like to offer three additional reasons why it appeared to be difficult for the teachers to remain focused on their perspectives of LA.

First and foremost, the second-phase curricular and didactic reform put the respondents and their learners under a great deal of pressure. We have already referred to the fact that, quite unforeseen, teachers were given more and often larger classes to teach, because the learners were expected to regulate their own learning. This meant that the actual contact time between the teachers and their classes was reduced. Reduction of time diametrically opposes commonly-held assumptions that reflection, and educational change for that matter, take up time. Besides, valuable time was spent on a host of school meetings that were not always perceived as useful by the respondents. In-between lessons and during school visits, learners

spontaneously told us about the pressures they felt, with more subjects to attend, and more tasks to complete. It forced many learners to become product-oriented in order to survive, whereas attention to LA requires a focus on the learning process.

Second, both the time and the opportunities were lacking to reflect on, investigate and discuss any of the dilemmas the teachers experienced in fostering autonomy in their learners. School managers heavily relied on the expertise of their teaching staff. The respondents had little time and opportunity to discuss dilemmas with their colleagues. If they did, it was usually geared at solving practical problems ad hoc. Opportunities for teacher education and schooling were either missing or were perceived as not very practical, helpful or relevant by the respondents. In 12.2 we will return to the contexts that appeared to be detrimental to change and development in more detail.

Finally, practical and challenging ideas that would enable the teachers to deal with dilemmas were needed, but were not on direct offer. That is why in section 12.3 we will plead for the creation of opportunities for teachers to reflect on their beliefs and practices by way of action or design research on one or more critical concepts related to the constructs of this study. Such research, which is carried out by the teachers themselves in cooperation with others, should first of all start with an exploration of what the teachers believe in and concentrates on critical links between their core beliefs and core academic theory.

This brings us to a discussion and analysis of a third gap we found between LA practice and LA theory. It concerns the finding that the respondents felt that few learners accepted responsibility for their own learning without teacher initiation and control. It leads us to discuss the issue of how feasible LA is as a pedagogical goal for adolescent learners.

The feasibility of autonomy for adolescent learners

We would like to discuss three reasons why, according to the respondents, only few learners appeared to accept responsibility for their own learning.

Our data first of all show that the teachers did not manage to have their learners understand and explicate what one is learning, why one is learning, how one is learning and with what degree of success in any structural way.

Secondly, our study shows that the learners were not or only marginally involved in mediating and negotiating important parameters of LA, such as determining the learning objectives, defining content and progression, selecting from useful strategies or techniques, or having the learners share in any decisions and initiatives taken.

Finally, any structural efforts by the respondents to have their learners plan, monitor and evaluate their own learning, both formatively and summatively, were largely missing.

Both our theoretical chapter on LA and our LA data have shown that autonomy is a complex and multi-facetted construct. It is action-based, transfer-oriented, and puts great demands on the knowledge and skills needed to carry out a task successfully, as well as on the type of tasks, actions and activities to be carried out. Its complexity is also shown by the fact that autonomous behaviour can neither be described easily, nor objectively. On top of all this, autonomy is not a steady state. It is dependent on many factors, situations and contexts.

In view of all this, the question seems justifiable whether autonomy is in any way feasible for adolescent learners who, often involuntarily, attend compulsory secondary education. How much more difficult it is to motivate adolescent language learners who study a language because they have to than it is to motivate more adult language learners who have chosen for a particular language course out of their own free will?

Yet, we would like to argue that autonomy is not only feasible for adolescent learners, it is even a *must* if we are to teach learners how to successfully learn to communicate in another language.

We feel the development of autonomy is not age-dependent. From the moment people are born, they develop and grow towards particular forms of autonomy. The degree and direction of autonomy may differ, even within certain age-groups. Nevertheless, broadly following Piaget, some general examples of autonomous development can be given. Babies soon learn they can manipulate their parents' behaviour by crying. Toddlers are often intuitively open to reasoning and the reasoning process. Primary-school children tend to develop a logic with the help of concrete referents. Thinking gradually starts to involve abstractions when children are between 12 and 15. Therefore, we feel human beings psychologically develop towards more autonomy, sometimes despite the fact that they are taught. Educators, parents and teachers alike, should initiate and facilitate this process.

We fully support our three respondents in their view that it is a fallacy to expect adolescent learners to be fully responsible for their own learning. Fostering autonomy is not simply a matter of having youngsters decide what they would like to do. We feel teachers should initiate learning. They should do so by creating a positive pedagogical and interpersonal learning environment. Learners should experience that a good sense of humour and social behaviour coincide with essential qualities such as discipline, perseverance and effort. Similarly, learners should sense that cognitive challenges are connected with the identification and development of the knowledge and skills required to successfully do a particular learning task or activity. Autonomy does not simply occur by having learners work on their own. Self-regulated *working* does not equal self-regulated *learning*, except perhaps for the happy few learners who prosper in any educational context or in any given situation.

In the next section, we will discuss and analyse how our teachers defined and attempted to operationalise the construct of communicative competence in their teaching and testing practices.

12.2.2 Communicative Language Education

Core theory

Three issues from chapter 4 have proved to be useful to interpret the teacher data on CLE. They were the concise methodological history of the teaching and learning of another language, Savignon's definition of communication and three models or specifications of communicative competence, and finally the ways in which form relates to meaning when languages are learned and used.

In chapter 4 we claimed that it is important for foreign language practitioners to learn from relative successes and failures of particular methodologies when other languages are taught and learned. We feel it is essential for teachers of another language to be aware of *what* has been taught, *how* this has been done, and *why* a foreign language has been taught in a particular way. It enables teachers to make justifiable choices for a particular method or didactic procedure and lead to, what we would call, healthy eclecticism. The one-and-only method for teaching and learning another language does not exist, despite the growing consensus that approaches should all be broadly communicative.

How, then, can communication and communicative competence be defined? In chapter 4, Savignon's definition of communication as a *continuous process of expression, interpretation, and negotiation of meaning* appeared to be helpful. The collaborative and constructive nature of this definition indicates that communication is about so much more than standard phrases such as 'Hello, I'm John. What's your name?'. In the theoretical chapters we also referred to three specifications of communicative competence, i.e. the models of Canale & Swain (1980), Bachman (1990) and the CEF, written under the auspices of the Council of Europe (1998 & 2001). All of these specifications derive from Canale & Swain's (1980) original model, in which communicative competence was seen to exist of four related competences: *grammatical*, *discourse*, *sociolinguistic* and *strategic*.

Perhaps most relevant to our discussion of communicate language education is the issue how attention to *form*, which aims at correct pronunciation, grammar or vocabulary, relates to a focus on *meaning*, which primarily aims at putting messages across. In other words should correctness be the main criterion of communicative ability, or should the intended interpretation of meaning be central to conceptions of communicative competence? Rather than favouring one over the other, we would like to suggest approaches that do justice to both, as recently Larsen-Freeman (2003) and Swain (2006) have done. Discussions of the relationship between meaning and form increasingly involve learner awareness and reflection. To add to this discussion, Larsen-Freeman coined the notion of *grammaring* as 'the process of doing grammar' and Swain the notions of *languaging* and *collaborative dialogue*.

CLE practice and CLE theory: potential links

We wish to highlight two potential links between the teacher data and the theory we discussed. First, all of the respondents share a view of language as communication and implicitly believe in communication as the expression and interpretation of meaning. Second, the respondents all aim at authenticity, albeit in different ways. The teachers believe in communication as the expression and interpretation of meaning. They all have their learners express themselves in English in written literary and/or project tests. The respondents attempt to stimulate creative thinking that results in views and opinions that are justified by the learners. Yet, the respondents' interpretations of how they have their learners achieve communicative ability differ.

According to Joy, CLE involves both the teacher's and the learner's ability to balance meaning and form. Neither drills nor a host of discrete grammar points should be focused on. In addition, she argued for more agreement on what is communicative and what is not. Joy's interpretation relates to what we have discussed in chapter 4, in which we attempted to define communicative competence in second and foreign language education. We have already mentioned that the issue of balancing meaning and form is a research interest that has recently been taken up again. We expect Joy, the budding professional, to be inspired by these concepts if she were given the opportunity to read about and discuss them.

When defining the construct of CLE, Mark primarily focused on his role as the teacher. His focus is solidly on the very discrete grammar points that do not very much appeal to Joy and Pete. By doing so, Mark has his learners regularly translate from English into Dutch and Dutch into English in order to have them develop awareness of contrastive form and usage. In chapter 11, we have already shown that Mark's approach interestingly shows remnants of the grammar-translation, direct and audiolingual methods and of a 'from surface structure to meaning' approach. Nevertheless, also Mark sees language as communication, speaks English a lot and has his learners speak and write English as well, albeit mostly on literary issues in our data. Mark more than lives up to his name as the literary master.

Pete stresses the notion of willingness once more in his interpretation of CLE. Communication is dependent on the willingness of a person to put meaningful messages across to an interlocutor who is willing to understand. Pete explicitly focuses on actual communication between two people. He wishes his learners to express opinions of their own on themes that are close to their hearts in English. Pete feels it is important for him to engage his learners, and for his learners to entertain, move and impress him.

A second link to the CLE theory we discussed is that the teachers all aim at authentic texts that focus on themes relevant to their learners' lives. Joy and Pete prefer making their own course materials. They do so by integrating a host of authentic texts and tasks in projects that are meant to engage the learners. Mark ultimately aims at dealing with a host of literary texts, which he has his learners read,

study and interpret. The first tests the respondents offered for discussion were all communicative writing tests.

CLE practice and CLE theory: crucial gaps

Based on our analysis of what was there and what was missing in our teacher data in relation to the theory we discussed in chapter 4, we would like to mention four matters. Once more, we mention the absence of explicit references to mainstream theory by our teacher respondents in this investigation. Secondly, we will go into the respondents' lack of specification of what communicative language ability entails. Then we will deal with the finding that our respondents appeared to have paid only limited attention to developing their learners' awareness of meaning, form and usage. Finally, we will deal with lack of attention to the view that communication requires negotiation.

CLE theory and CLE practices

Despite the links we found between the respondents' teaching and testing practices and the theory we outlined, we found no explicit evidence that any of their practical theories was based on knowledge of some depth of the issues we discussed in chapter 4.

We would like to advocate that particularly the knowledge of models of communicative competence and the specific knowledge and abilities that underlie the construct need to be focused on. We have already hinted at the potential of Larsen-Freeman's concept of *grammaring* and Swain's notions of *languaging* and *collaborative dialogue* in this respect.

We would like to argue that theory might help to address what we found to be missing in our teacher data on CLE, i.e. a lack of specification of what communicative competence involves and the development of their learners' awareness of meaning, form and usage.

Lack of specification of what communicative language ability entails

Even though the respondents all share a global view of language as communication, their specifications of underlying knowledge and skills of communicative competence rarely went any further than mentioning essential grammar points and, to a lesser extent, some references to vocabulary.

In the investigation, the respondents did not appear to be familiar with important knowledge and particular skills and competences that are generally seen to underlie communicative language ability. Within a particular competence, such as linguistic competence, the respondents seemed to focus more on knowledge of grammar than on phonological, lexical or semantic competences. Within their grammatical knowledge, the respondents do not distinguish between declarative (knowing what) and procedural (knowing how to) knowledge. Occasionally, knowledge and skills were even mixed up. Neither did our data reveal any serious attention being paid to sociolinguistic, pragmatic or strategic competences. The data did not disclose much attention to markers of social relations, register differences or dialects and accents. Of the pragmatic competences, only discourse competence received some attention, albeit primarily grammatical, with a focus on word order and the ways in which paragraphs relate to one another. There were no convincing data on pragmatic competences, such as functional competence or schematic design competence. Likewise, reception, production, interaction and mediation strategies received little or no attention.

As a result, the respondents were unable to specify communicative objectives in any detail. If communicative objectives were mentioned, they were closely related to their beliefs. Mark felt that communication about literature allowed learners to learn about English and about life, once a solid grammatical and idiomatic foundation had been laid. Pete and Joy believed that engaging projects and doing a host of related

practice tasks in English would ultimately lead to learners expressing themselves in English.

Limited attention to developing the learners' awareness of meaning, form and usage

When the learners were being prepared to the communicative writing tests that featured in this investigation, only limited attention was paid to having the learners reflect on the ways in which form and usage relate to meaning. Joy and Pete's focus on having their learners reproduce the correct verbal tense and form of an authentic English text, impressive though it was, was basically procedural. It involved the application of a heuristic verb scheme, and only implicitly focused on the meaning of the text. When Mark prepared his learners for the UF test, which was based on the regular course materials, the focus seemed to be on the products rather than on the reflective processes that lead up to correct response, with the learners largely working on their own.

Little or no attention to the fact that communication requires the negotiation of meaning

We have already referred to the respondents sharing the view that communication requires the expression and interpretation of meaning. The teacher data illustrated attention both to the expression and to the interpretation of meaning. Yet, no evidence was found of an explicit focus on the fact that in the process of communicating written or oral messages, meaning often depends on the ways it is negotiated between the interlocutors or the writers and their expected readers.

In view of the negotiation of meaning, we feel it is a serious drawback that the writing tests in this study were hardly ever evaluated with the learners in any detail. Learners were generally not given any feedback other than marks or some remarks on the particular ways in which they had attempted to express, interpret, and, above all, had tried to negotiate meaning in the tests they did.

Analysis

In our analysis, we will first of all discuss reasons why we found what we had found regarding the four issues presented above. In addition, we will go into two findings related to communicative language education that were prominent in this study, but could not be related to the contents of the theoretical chapter in which communicative competence in foreign language education was discussed. These additional findings were the prominent position of literary texts and the role of published course materials in teaching the learners how to communicate in English.

A restricted view of communicative competence

Similar to our findings with LA, the respondents never directly referred to academic theory or to any key publications related to the construct of communicative competence in foreign language education. It resulted in restricted interpretations of communicative ability, which were mainly inspired by the respondents' core beliefs.

We will not repeat our earlier analysis here of why there were gaps between what the respondents believed and did and their teaching and testing practices of CLE. The reasons we offered in our discussion of LA also partly apply to our finding that the respondents have a restricted view of communicative competence.

However, we would like to argue that the gap between theory and practice for conceptions related to communicative language education is the more remarkable. After all, communicative objectives have been formulated and communicative course materials designed in the Netherlands since the 1980s.

Time cannot be used as a factor here. Neither can the pressures of the second phase reform. For some twenty-odd years, FL-teachers have been given the opportunity to learn how to teach their learners to communicate in English and have their learners experience the ins and outs of it. Yet, again, only some of the key

notions of communicative competence appear to have filtered down to our respondents over the years.

Communicative language ability is a construct that is at least as complex and multi-facetted as LA has turned out to be. Its complexity and any of the underlying knowledge and skills other than the ones related to grammatical competence were paid little attention to. Perhaps the implicit assumption with communicative competence has been that by making the course materials more communicative, the teachers could not but follow and make their lessons more communicative. Our case studies suggest that the use of more communicative materials in the classroom does not simply lead to learners becoming better users of English.

We would like to contend that first-hand knowledge of models of communicative competence (Canale & Swain,1980 ; Canale, 1983; Bachman, 1990) and their broad specifications in the CEF (1998, 2001) might be particularly helpful in having teachers develop a wider perspective on communicative competence and the knowledge and skills it is supposed to entail.

Learner awareness of meaning, form and usage

In our discussion of the findings related to the construct of LA, we have already seen that the respondents did not or only marginally focus on metacognitive and metalinguistic reflection and on formative assessment and evaluation.

We see reflection and evaluation as essential components in having learners develop awareness of the relationships between form aspects of the language, knowledge of usage characteristics and meaning as it is expressed, interpreted and negotiated between two or more people.

The question here is how teachers can possibly stimulate learner reflection and the learner's self-assessment and evaluation skills, if the teachers themselves have not developed a propensity towards reflectivity and are unable to specify the knowledge and skills required for successful task performance.

We would like to argue that it is easier for teachers to make learners aware of matters they themselves have acquired a broad knowledge of and have had ample experience with. A teacher of another language first of all has to be a proficient user of the foreign language and in addition is able to clearly indicate and explain what declarative knowledge, procedural knowledge, specific skills and competences allegedly are involved when learners learn how to communicate in that language. Such teachers are likely to be more capable of making learners aware of meaning, form and usage than educators who can only generally indicate what communication or communicative competence is. We therefore consider broad academic knowledge of central concepts of foreign language teaching, such as communicative competence, a must for teaching professionals in the field.

This does not mean that we expect language teachers to successfully address any issue related to communicative language education by simply learning more about communicative ability. Communicative language education is still an area that is widely investigated. A lot of what is known in the world of academe about language teaching and learning is still based on particular interpretations of central constructs or on hypotheses about effective teaching and learning strategies.

Yet, we do feel that effective teachers are not only capable of creating challenging and safe learning environments, but are also informed about what is known about main constructs of the subjects they teach. Communicative competence is such a main construct for language teachers.

Communication requires negotiation

We would like to argue that if communicative tests are not evaluated with and by the learners, the writing the learners produce will basically require the expression and interpretation of meaning as practised in the preparatory tasks and activities in class. We surmise that more than a single learner will therefore safely reproduce the ways

in which meaning was interpreted by their teacher in the course of the tasks preparatory to the communicative writing test.

A focus on products rather than on the process and a focus on reproduction rather than on creative thought neither seems to stimulate the development of autonomy, nor the cultivation of communicative ability.

If anything, this study has highlighted that it is essential for teachers and learners to discuss what is meaningful and what is not in communicative language education.

We will finish our discussion of communicative language education with two findings that could not be directly related to the contents of our theoretical chapters.

The role of published course materials

Interestingly, we have seen that Mark largely relies on the selection and gradation of course content in the text- and workbooks that are used at his grammar school. This means that Mark implicitly adheres to and follows the specific ways in which communicative ability is interpreted in these materials. In Mark's case it has not always led to lessons that can be called communicative.

Pete and Joy started out with using published course materials at the beginning of the school year. They, however, soon concluded that the projects, tasks and activities that had been used before were considered more successful in teaching their learners to communicate than the regular course materials appeared to do.

This finding first of all strongly suggests that the course materials teachers opt for and use have to correspond to some extent with their ideas of effective teaching and language learning. If this is not or no longer the case, the materials tend to be replaced by others.

Secondly, if course materials positively relate to a teacher's beliefs regarding effective teaching and testing and at the same time are based on a more extended conception of communicative competence, the effect may be that the teachers' interpretations of communicative language ability is also extended. That way, course materials may become agents of professional development.

Thus, by influencing what teachers do, course materials might affect what teachers think. However, we sense that affecting teachers' beliefs by merely having them experience new didactic procedures they have never used is likely to be as formidable a task as directly attempting to influence how teachers think. Nevertheless, the exact ways in which communicative course materials are used by teachers and learners invites further research.

There is a second finding that could not be directly related to our theoretical chapter on communicative competence.

The use of literary texts

All of our respondents enjoyed incorporating and dealing with literary texts in the materials they used. It strongly suggests that poems, challenging song or rap texts, short stories, plays, movie scripts and full novels potentially provide challenging materials that help learners develop their communicative skills. In the learning and teaching of foreign languages, poems, short stories and the reading of full novels used to have a fairly prominent position in the curriculum. In the second phase, this position was initially largely marginalised, much to Mark's regret in particular.

The data of this study have, at times convincingly, shown that the use of literary texts engages the teachers and allegedly their learners as well, even though Mark approaches literature from a more cognitively-oriented angle than Pete and Joy do.

Given the fact that literary texts often inherently produce the best possible use of language one can find on themes that often relate to the lives of human beings, we would very much like to advocate the use of a variety of literary texts as incentives to have learners communicate in that language.

More often than not, literary education is seen as a subject with objectives of its own that are unrelated to the process of language learning. Even though we realise there are valid arguments for using the first language when foreign language literature is dealt with, we would like to plead for a more prominent role of target language literary texts when learners learn to communicate in another language. We feel it is too easily assumed that gaps in linguistic command invariably cause learners to lose interest in FL literary texts. The successful exploration of literary texts by our respondents hints at the opportunities that such FL texts have in foreign language learning.

However, in view of the development of the learners' autonomy and skills as readers, it is important that an informed selection of literary texts is made, with opportunities for choice for the learners who are learning how to read, interpret and communicate about the issues addressed in a particular literary text.

We have now arrived at a discussion and analysis of our findings related to the third construct central to this study.

12.2.3 Foreign Language Assessment and Evaluation

Core theory

Five issues from our theoretical chapter 5 on foreign language assessment and evaluation particularly added to our understanding of the teacher data.

There was first of all the need we felt to distinguish more clearly between testing, assessment and evaluation. Both in the literature and in our data on language testing distinctions between the three were not always clear. We particularly highlighted the importance of evaluation in language testing, which we defined as the retro- and prospective inferences made from the results of a particular assessment procedure. We consider evaluation as an important link between LA and CLE, with its focus on definable and observable knowledge, skills or particular competences and its relationship with metacognitive and metalinguistic reflection.

Secondly, in all of our theoretical chapters we have approached the constructs from a socio-historical perspective. Following Spolsky (1978), this led to our discussion of four trends in the field of language testing, respectively the *pre-scientific*, *psychometric-structuralist*, *integrative-sociolinguistic*, and *critical-dynamic* trends. We feel that knowledge of these trends may help teachers identify and reflect on characteristics of their own assessment and evaluation practices in relation to basic tenets of the four directions.

Thirdly, we discussed the relationship between test purpose and test type. We would like to argue that knowledge of and reflection on this relationship will potentially help teachers to make their assessment procedures more valid.

Fourthly, there were two essential measurement qualities of tests we singled out for discussion, i.e. *reliability* and *validity*. We particularly feel that knowledge of and experience with the notion of construct validity may be an important impetus for improving the quality of teacher-made tests. We feel that the construct validity of teacher-made tests is likely to be enhanced if teachers reflect on the validity of the inferences they make on the basis of particular test results, together with colleagues, their learners and knowledgeable professionals in the field.

Finally, the essential measurement qualities of reliability and construct validity appeared to have been embedded in the notion of test usefulness developed by Bachman & Palmer (1996), who defined the notion as the sum of six interrelated qualities, i.e. reliability, construct validity, authenticity, interactiveness, impact and practicality.

Testing practice and Testing theory: potential links

There are two potential links between the teacher data on foreign language assessment and evaluation and the theory we discussed. The first is the acknowledgement of the three respondents that marks generally are important for their learners. The second link is the basic concerns that the respondents exhibit with the essential measurement qualities of reliability and validity.

All of the respondents have expressed that marks are important for the majority of their learners. Learners are very much aware that marks of 6s or more on the Dutch 1-10 scale allow them to move on to the next form and ultimately pass their final exams.

In addition, the respondent teachers state that marks can encourage as well as discourage learners. Particularly Pete and Joy are of opinion that insufficient marks, or marks that are considered bad by the learners, tend to discourage rather than to stimulate them. Mark seems less worried about the discouraging effects of marks, and stresses that tests are necessary incentives for his grammar-school learners to learn. It would seem that gifted learners are more likely to be challenged if they do not get the mark they are used to getting or feel entitled to get.

We would like to conclude that by acknowledging that marks are important for learners the respondents implicitly acknowledge the potential power of tests to affect the ways in which the learners learn.

Yet, our teacher data has shown little or no evidence of any ways in which our informants have exploited this potential power of tests. Washback of assessment procedures generally was not on their minds. We therefore feel that washback effects always exist, but that these effects will remain very much implicit unless they are explicitly addressed by the teachers and the learners.

The second potential link between the testing data and core testing theory is that the respondents show some basic concerns with scoring and scorer reliability. All of the three respondents feel it is important to score in consistent ways, which goes both for the ways in which they score work of different classes and for the consistency of scores and scoring between themselves and their direct colleagues.

The respondents' construct interpretations of effective testing also implicitly incorporate the notions of transfer of important knowledge and skills and, basically, content validity. They all feel that the knowledge and skills that have been practised by the learners are adequately and representatively assessed in a summative test.

We would like to argue that it is easier to interest teachers for more extended reliability and validity concerns if such theoretical insights link up with what they consider valuable and essential in their assessment and evaluation practices.

Testing practice and Testing theory: crucial gaps

We feel six of the findings of our teacher data on testing need to be addressed in particular. The first issue is similar to the ones we highlighted in our discussion of LA and CLE. Despite, or should we say because of, the abundance of specialist literature in the field of language assessment and evaluation, the respondents did not refer in any way to testing theory. A second finding that calls for attention is that our respondents test and assess, but generally do not evaluate their assessment procedures with their learners. Thirdly, there is the finding that the respondents all have a summative view of language testing, and neither assess nor evaluate formatively. A fourth aspects of our data that calls for discussion is the fact that the respondents focus on achievement rather than on proficiency testing. Fifthly, our respondents experienced difficulties in specifying the knowledge and skills they felt to underlie complex skills such as communicative writing ability. Finally, whereas aspects of the pre-scientific, psychometric-structuralist and integrative sociolinguistic trends in language testing could be traced, this was not the case for the more recent trend we labelled as critical-dynamic.

Analysis

Testing theory and testing practices

The rift we found between LA theory and LA practice was big, the one between CLE theory and CLE practice bigger, but it was biggest between academic theory on testing and the respondents' assessment and evaluation practices. None of the informants made any direct references to testing theory. Not even Joy, who was among the more successful students at the graduate school of education that she had just left.

In our previous discussion of the gaps between theory and practice, we have already suggested a number of reasons why the gaps appeared as they did. In this section, we will discuss some additional reasons that somehow seemed to have made background literature and central concepts in the field of language testing even less accessible to the respondents than theoretical knowledge and research results in the fields of LA and CLE.

First of all, we feel this might be caused by the overwhelming abundance of literature and research in the field of language assessment and evaluation. Studying and reviewing the literature does not only take up time, it also requires more specialist knowledge than the other two constructs seem to require.

Secondly, all of the respondents primarily developed their assessment and evaluation practices on the basis of practical experiences with tests as learners and early users and constructors. Once their preferred ways of testing and tests were opted for early in their careers, their testing practices tended to become stable and immovable, not unlike their core values and beliefs. Likewise, the informants' assessment and evaluation procedures remained stable throughout the year of data collection and did not seem to be affected by the critical questions that were repeatedly asked when the three sample tests were discussed throughout the school year.

Even though our informants did come up with interesting, impressive and illustrative tests, we feel this lack of reflectivity regarding their testing practices calls for serious attention. The issue is particularly relevant, because we feel a focus on LA and CLE requires a less static, more dynamic and more informed approach to language testing than the ways in which assessment and evaluation has been interpreted by the three teacher respondents. This issue is also important in view of the power that tests, scores and marks invariably have in determining success or failure in education.

We would like to argue that evaluating tests and their results with peers and the learners themselves are necessary steps to improve the quality of teacher-made tests and create positive washback of these assessment procedures on the ways in which the teachers teach and the learners learn how to communicate in English.

The argument used most often by the respondents why they did not evaluate tests and test results with the learners in some detail was lack of time. As a number of summative tests had been laid down in the so-called plans for testing and summative assessment (PTAs), the respondents felt that no valuable time could be lost to concentrate on bygones. New tests would create new opportunities for the learners to score. The fact that language learning invariably concerns cyclical processes did not receive much attention in the testing practices of our respondents. This particular finding leads to a second issue we would like to discuss.

The respondents basically see tests as summative assessment procedures. The respondents seem to have accepted this as a given fact. Testing is seen as a summative procedure at the end of a particular learning process. Only Mark specifically mentioned that he regretted the lack of flexibility with testing in the renewed second phase, due to the fact that tests had been laid down by contents and date at the beginning of the school year.

Despite the PTAs, we would like to advocate a broader perspective on language testing that incorporates formative assessment and evaluation procedures from the beginning of the learning process onwards. This does not mean that we argue for an increase in the number of informal tests the learners get. It is just that educational practitioners tend to forget that asking particular test questions or having the learners do a brief test task or activity of a couple of minutes, perhaps first individually and then in pairs, is and should be very much part of effective assessment and evaluation practices. All of the test questions and in-between test tasks and activities should justifiably and observably relate to the contents of the summative test that is ahead and was already planned at the beginning of the year.

In addition, we feel that learners should be taught to assess and evaluate themselves and their peers. Self-assessment and self-evaluation will be particularly effective when performed in the course of preparing for a particular summative test.

We also hypothesise that if teachers and learners focus more on formative assessment and evaluation, the summative tests and their results are more likely to have positive washback effects on the ways teachers teach and learners learn. Above all, formative assessment and evaluation procedures create opportunities for teachers and learners alike to change or adapt the learning processes that are seen as preparatory for a particular summative test or exam to come. We consider the absence of formative assessment and evaluation procedures among the more important findings of this study.

A finding of a different nature than the first two mentioned above is that we found no evidence of proficiency testing in our testing data. A proficiency test or assessment procedure is a test of the learners' language ability without having them specifically study for that test beforehand. Proficiency tests often tend to be standardised and criterion-referenced, but can also be constructed by teachers, who confront their learners with assessments of their proficiency without having asked them to prepare for these tests in any way.

By contrast, the tests in this study were all achievement tests, that is they were tests that directly related to the learning process and activities that went ahead and that therefore did not tap directly into the learners' overall or particular language ability at a given time.

We feel that particularly the use of more standardised, criterion-related proficiency tests at schools will be beneficial. Three boons can be mentioned.

First, it will allow both the teacher and the learners to assess language ability at any given moment in the learning process, providing reliable and valid information on what learners know about or are able to do in another language. Second, it allows for a variety of reliable comparisons, e.g. between learners in one class, learners in different classes, and learners in different schools. Third, the scores the learners get are more likely to motivate teachers and learners to adapt the learning process than teacher-made tests tend to do, because the latter often are dependent on the reproduction of particular knowledge and skills that have been practised in the course of the learning process.

At many schools in the Netherlands, there is increasing attention to practise materials and exams that lead to external certificates of English proficiency, such as the Cambridge or Anglia exams. We welcome such alternative and standardised assessments of general proficiency. They will particularly be effective if results are assessed and evaluated formatively. There is yet another matter that calls for discussion.

In the course of the year of data collection, it appeared that particularly Joy and Pete experienced difficulties in specifying the knowledge, skills and understandings they felt to underlie complex language skills such as writing ability. Mark seemed to experience considerably fewer difficulties in specifying the knowledge and skills that underlie a particular test task. However, we should in this respect add that Mark focused on a host of grammar points, the reproduction of idioms or the replication of

particular literary interpretations. Pete and Joy generally object to what they see as mindless reproduction.

Nevertheless, we would like to argue that if teachers do not know what exactly needs be learned and practised in order to teach learners how to write more effectively and correctly in English, the only thing left for learners is to practise and practise a lot in the hope that some day this very practice makes perfect. Such an approach of learning by doing may work for a limited number of intelligent and gifted learners. However, we have serious doubts about its beneficial effects for less gifted and less motivated learners.

What alternative can then be offered? Although we do not at all object to giving learners time to practise, we would like to plead for more teacher direction and support in the ways in which the learners practise their writing or attempt to improve their overall communicative abilities in another language. Our plea, therefore, is for interchanging and balancing *quality time*, in which the teacher directs , explains, specifies and above all motivates with *practice time*, in which learners first individually and then in groups, experience first-hand, practise and subsequently improve what has been dealt with in class. It brings us to a final matter we would like to mention.

In chapter 5, we discussed the critical-dynamic trend as the last of the four trends in foreign and second language testing. Some of the challenging and relevant issues we explored in our discussion of this trend were collaborative dialogue, the mutual responsibility of test giver and test taker, the rights of the test taker, forms of dynamic assessment and the ways in which dynamic assessment relates to interventionist and interactionist approaches and to planned, incidental, internal or external forms of formative assessment and evaluation.

None of the data showed any attention to the issues associated with this trend, perhaps with the exception of the teachers objecting to the relative importance and power of the national reading examination tests havo- and vwo-learners have to sit at the end of their school careers, the results of which decide for 50% the final mark for foreign languages the learners are going to get.

The respondents were generally felt to have a love-hate relationship with the national reading examination tests. On the one hand, they accepted and welcomed the reliability of the scores on the tests and the structure it gave to their teaching practices, but on the other they objected to the sheer weight that was given to the assessment of reading skills over the knowledge and the skills they felt to be more important for their learners to be tested on. The respondents also frequently questioned the validity of the national reading examinations and claimed that it favoured learners with analytical minds and a broad knowledge of the world. Thus, the tests were felt to assess more than English language ability as they had defined it.

12.3 A context of innovation and change

In this section we will return to the research question in how far the second-phase reform has been conducive to fostering learner autonomy, enhancing communicative language ability, and promoting effective assessment and evaluation.

In three parts, we will go into what we found, analyse why we found what we had found, and finally highlight three reasons why the context of innovation and change did not live up to the expectations in the year of data collection. In our theoretical chapters we have attempted to consistently investigate what we can learn from the past. We suggest anyone involved in new surges of innovation and change do the same.

12.3.1 What we found

Our teacher data has shown that the second-phase curricular and didactic reform has only been successful for the construct of LA to a limited degree. What the reform at it is best achieved was that it had our respondents reflect on LA and come up with interesting perspectives on the construct. It also had some of their learners successfully participate in group work, which is proof of the opportunities of cooperative learning.

Yet, reflecting on LA perspectives and putting them to the test in everyday practice was either problematic or was found very little evidence of. As far as enhancing communicative language ability and promoting effective assessment and evaluation were concerned, we found no convincing evidence of any changes in the respondents' practices from what they had been used to doing from the start of their careers.

This by no means disqualifies our respondents. By keeping up what they believed in and doing what they were considered to be good at, they managed to pull themselves and their learners through a year of turbulent curricular and didactic change. The respondents positively surprised us with a number of impressive projects and tests, which were illustrative of their beliefs. So, is our conclusion then to leave well alone and just let our teachers be? No, it is not. For each of the central constructs to this investigation we found ample, and we would like to argue, necessary room for professional development.

12.3.2 Why we found what we have found

We feel there are at least six reasons why the second phase innovations only led to limited changes in the respondents' teaching and testing practices.

First, there was the implicit assumption that didactic innovations could be introduced top-down, with little consideration for the beliefs and qualities of practising teachers. We have already argued how important we feel it is to take teachers' beliefs as a starting point of professional development.

Second, the second-phase reform involved innovations of both the curriculum and at the same time expected teachers to fundamentally change the didactic procedures they had come to rely on. This caused frustration and a host of practical problems at the schools where the three respondents taught. It required a lot of time, energy and hard work to remedy these problems by the school managers, the teachers and their learners. The result for our respondents was that there was only limited time to reflect on or experiment with the innovations they were expected to implement.

Third, the theoretical background of the proposed didactic changes, which could be characterised as broadly constructivist, did not receive a lot of attention. Besides, there was only limited empirical evidence available that LA principles could be implemented with success in regular upper secondary education.

Fourth, when the second phase was introduced, it appeared that teachers were generally given more and larger classes to teach, as the learners were expected to regulate their own learning. Thus, cost-effectiveness and efficiency became more important than the quality of education. Ironically, the reduction of teaching time has somewhat slowed down the effects of the massive shortage of academically-trained teachers in the Netherlands, which had been forecasted by specialists for over a decade.

Fifth, there was the fallacy that by having adolescents work on their own, they also successfully learn on their own. There is no one-to-one relationship between the two. We have already argued that, together with our respondents, we feel that teacher initiation and control are decisive factors in implementing experiencing any principles of LA. We have also argued that formative assessment and evaluation and metacognitive and metalinguistic reflection are required to have learners regulate their own learning with some promise of success.

Sixth, the effects on the learners of an overloaded curriculum and the lack of subject-specific supervision during practice tended to make learners even more product-oriented than they used to be. We also found that having the learners work on their own did not guarantee that they learned how to use English appropriately and correctly. If such a finding is consistent with others, this may potentially lead to a loss of what adolescents know about and are able to do in English.

This was also sensed by our respondents. Mark felt that more attention had to be paid to grammar and idioms with his fourth formers than he was used to. Joy and Pete found that having their learners construct and build idiom files of their own resulted in the learners understanding and using fewer idioms than they used to.

We realise that particularly the last reason we offered why the second-phase innovations cannot be considered successful is a bold one, given the scope and limitations of the present study. It is impossible to generalise our results other than to theory or to other studies. After all, our data has been retrieved from only three respondents, who cannot be called representative in any way. We have already related our findings to what we have called mainstream theory. In addition, we feel the need to briefly relate our findings to two studies that were carried out and published while the present study was written. One is an evaluative investigation carried out under the auspices of the process management second phase entitled *1998-2005: Zeven Jaar Tweede Fase: Een Balans* (2005). The other is a study entitled *The Assessment of Pupils' Skills in English in Eight European Countries 2002* (Alabau, I., Bonnet, G, de Bot, K., Bramsby, J., Dauphin, L., Erickson, G., et al., 2004). It was a project commissioned by the European network of policy makers for the evaluation of education systems. The findings of these studies seem to corroborate that the level of English of adolescent havo and vwo-learners in the Netherlands is a cause for concern. Mathematics and English were mentioned most often as subjects that did not link up well with what was expected of first-year students in tertiary education.

Second-phase havo and vwo graduates in higher education

The study *Zeven Jaar Tweede Fase: Een Balans* (Tweede fase Loket, 2005) evaluated the ways in which the knowledge, insight and skills of havo and vwo graduates linked up with what was expected of them at colleges or universities for learners who graduated from 1998 to 2005. On the positive side, the study showed that both the students and their college or university lecturers felt the students were competent in general skills, such as abilities related to planning and presentation. However, the students as well as their lecturers in higher education were dissatisfied with the students' language ability, arithmetical skills, accuracy, and analytical skills. For mathematics this particularly concerned the sectors technique, economy, nature and behaviour and society. For English it concerned all sectors, with a difference in perception between vwo and havo graduates. Vwo graduates felt too much English knowledge and skills were expected of them ,whereas havo graduates feel too little was expected of what they knew about or could do in English. In addition, the graduates themselves indicated they would have liked their secondary schools to have paid more attention to communication skills, competences related to cooperation and taking initiatives (havo), and to study planning and analytical skills (vwo).

The study corroborates our findings that more attention needs to be paid to the learners' communicative and analytical skills in the teaching and testing practices we studied. We feel far more English has to be used by the learners during practice. In addition, we are of opinion that a focus on metacognition, metalinguistic reflection and on formative assessment and evaluation when learners learn to communicate in another language will help learners to improve both their reasoning and analytical skills. There is much to be done and won in these areas.

English skills of Dutch secondary-school graduates compared with European peers

The study entitled *The Assessment of Pupils' Skills in English in Eight European Countries 2002* (2004) involved a large-scale comparison of the English skills of secondary-school graduates in eight European countries. We would like to present the scores of the graduates on an assessment procedure that tested for linguistic knowledge (25 items), written comprehension (16 items), written expression (21 items), and oral comprehension (13 items). We will present the results for seven countries in four scales. The results for Germany have not been included, because data retrieval in Germany was different from the way it was retrieved in the other seven countries. In each scale, the second column lists the scores, and the third column the standard deviations.

Oral comprehension scale:

1.	Norway	73.26	19.60
2.	Sweden	72.18	19.65
3.	Denmark	64.77	20.07
4.	Netherlands	61.63	21.44
5.	Finland	59.65	24.52
6.	Spain	38.33	23.08
7.	France	30.60	20.39

Linguistic competence scale:

1.	Finland	67.59	22.10
2.	Norway	66.36	20.40
3.	Netherlands	65.00	22.00
4.	Sweden	64.23	20.43
5.	Spain	58.75	23.30
6.	Denmark	53.95	22.10
7.	France	48.01	21.41

Reading comprehension scale:

1.	Sweden	85.88	22.31
2.	Norway	82.03	26.82
3.	Finland	80.29	23.07
4.	Denmark	78.32	26.26
5.	Netherlands	77.47	21.54
6.	Spain	63.57	21.66
7.	France	56.84	21.85

Written production scale:

1.	Norway	56.30	29.69
2.	Sweden	55.39	28.04
3.	Finland	47.70	29.46
4.	Denmark	46.17	29.33
5.	Netherlands	46.04	25.67
6.	Spain	23.41	25.50
7.	France	14.55	17.81

What should our conclusions here be? The more one travels to the north in Europe, the better English is understood and written? The darker the days in winter time, the better the level of English is?

On a more serious note, given the fact that the Netherlands is a small country, with English being offered as a school subject for all learners from group 7 in primary education onwards, with English often used as a lingua franca by many of its inhabitants, with a lot of English exposure on TV and via the Internet, we feel the results for the Dutch second-phase graduates are disappointing as compared with their Scandinavian peers. The scores are particularly disappointing for written production.

The question seems justified to ask ourselves whether we should be pleased with the results of our havo and vwo graduates in view of the knowledge of and skills in English we teach at secondary schools in the Netherlands. Again, we conclude there is something to be done and won.

In the next section, we will discuss what can be learned from the investigation in view of further research and educational programmes on how to create positive washback of assessment and evaluation practices on classroom teaching and learning.

12.4 Recommendations for pre- and in-service teacher education

In our suggestions on how to create positive washback of assessment and evaluation practices on foreign language teaching and learning, we do not distinguish between pre-service and in-service teacher education. We feel our advice concerns both the practices and education of beginning teachers, who have not yet qualified, and of more experienced teachers, whose beliefs and construct interpretations tend to be firmly rooted in their daily practices.

There is one main reason why we do not distinguish between pre- and in-service teaching. We feel teacher education and reflective teaching practice are dialectically intertwined in their mutual attempts to relate theory to practice, and practice to theory.

Teacher education in the Netherlands is geared at educational practice from the beginning onwards. This is increasingly the case for primary and secondary teacher education, both for degree-one and degree-two teachers. We welcome and support this development. You cannot learn how to teach only from books and laboratory-like experiences at institutions for teacher education. In addition, we feel that there is no guarantee that beginning teachers will develop into professional teachers if they are merely coached by experienced teachers. Student teachers soon start with their probationary work in practice. Here, beginning teachers are more than once confronted with remarks to forget about academic theory and instead heed the practical advice they get from experienced teachers. This way undesirable rifts may occur between theory and practice, with practical advice being perceived as unacademic, and academic theory as unpractical.

Teacher education in the Netherlands is geared at educational practice from the beginning onwards. This is increasingly the case for primary and secondary teacher education, both for degree-one and degree-two teachers. We welcome and support this development. You cannot learn how to teach only from books and laboratory-like experiences at institutions for teacher education. In addition, we feel that there is no guarantee that beginning teachers will develop into professional teachers if they are merely coached by experienced teachers. Student teachers soon start with their probationary work in practice. There, beginning teachers are more than once confronted with remarks to forget about academic theory and instead heed the practical advice they get from experienced teachers. This way undesirable rifts may

occur between theory and practice, with practical advice being perceived as unacademic, and academic theory as unpractical.

We therefore feel that teacher educators, in-service teachers and pre-service teachers should be willing and able to bridge any of the perceived gaps. It requires that teacher educators, in-service teachers, and pre-service teachers accept that they are life-long learners, who learn from collaborative dialogue and the specific and essential input they all have in the teaching and learning process. However, achieving all this will neither be easy, nor cost-effective. Nevertheless, we feel it is essential that time, money and energy will be spent on attempts to realise what we suggest.

We will outline our recommendations in four sections. In the four recommendations, a central role has been assigned to the cooperation of educators and researchers in reflective action or in educational design research. We feel that attention to these two research paradigms holds more promise of success for educational change and professional development than top-down efforts of educational reform and the introduction of new course materials so far have had.

12.4.1 Starting with teachers' core beliefs and domain-specific construct interpretations

This study has provided insight into the core beliefs and domain-specific construct interpretations of three academically-trained good-practice teachers of English in the Netherlands. We have provided evidence of the importance of these beliefs and interpretations as they pervaded the teaching and testing practices of our respondents and turned them into the teachers they were at the time of the investigation.

We are convinced that professional development should start with allowing and enabling teachers to express, interpret and negotiate notions that relate to what they consider to be effective teaching, learning and testing. What teachers really think and feel often remains hidden or is implicit in their teaching and testing practices.

Having educators express, interpret and negotiate beliefs and important domain-specific notions and concepts is far from self-evident. When we piloted the interview guides that have been used in this investigation with a lifetime friend of ours, we touched upon beliefs and notions we had never, ever discussed in any detail. People generally like expressing what they believe in most.

Once it has been established what teachers feel, think and believe in, attempts should be made to motivate and facilitate them to critically reflect on their current teaching and testing practices. Therefore, as a next step we feel it is essential to motivate and facilitate teachers to critically reflect on their teaching and testing practices.

12.4.2 Priming teachers for lifelong learning and professional development

First and foremost, teachers have to be facilitated in time. We realise that time is money, especially in societies where individual gain is considered more important than collective well-being. This has too often resulted in the lack of investments in education and other matters that concern the well-being and future of all members of society, such as health care and environmental concerns. Politicians and managers too often tend to think in the short term, which is likely to lead to future problems that could and should have been addressed. We would like to contend that unless teachers are facilitated in time, the recommendations of the present study can be considered as not being written. We cannot and should not shut our eyes and ears to the fact that in order for teachers to perceive their work as a profession instead of a job, they need to be given fewer hours to teach to smaller classes with better career prospects for teachers who are willing and able to develop as professionals. In addition, teachers who are willing and able to develop should be given opportunities investigate their own practices in relation to core theory. In this investigation, we have

highlighted instances of core theory and related these to the teachers' beliefs and teaching and testing practices. Yet, merely facilitating teachers is not enough to have teachers critically reflect on their educational practices. A second issue needs to be addressed. It concerns a teachers' own willingness and ability to develop as professionals.

Earlier in this study, we referred to Van den Akker & Bergen warning us that many attempts at educational innovation have stranded on 'the stubborn rocks of habitual change' (1997: 121). It is not easy to change teachers' beliefs and challenge them to develop as professionals. That is why we feel teachers have to be motivated for lifelong learning both intrinsically and extrinsically. Motivating teachers is not unlike motivating learners. Motivation starts by building on the specific expertise that teachers have and encouraging them to address matters that concern them most. Yet, merely motivating teachers intrinsically will not be enough. In addition we feel that extrinsic incentives, such as having the teachers apply for fewer classes to teach in order to carry out action or design research or offering them extra pay if they successfully participate in projects, will decisively add to their motivation to learn and further develop their professionalism.

12.4.3 Relating teachers' beliefs to core theory

Once teachers have been facilitated and motivated to critically explore their beliefs and practices, we feel it is useful to introduce key articles and background literature that link up with what they believe in, already know or have already done.

In this study, we have discussed instances of core theory that are likely to be explored with success by the respondent teachers. Our theoretical explorations in chapters 3, 4, and 5, have led to aspects and notions that refer to the constructs of LA, communicative competence, and foreign language assessment and evaluation. Our analyses of the teacher data and the ways in which we related them to core theory have suggested areas in which a lot has to be done and won if teachers and researchers concentrate on them.

We feel the ways in which our respondents interpreted LA, CLE and foreign language assessment and evaluation should have been more informed by domain-specific knowledge of the three constructs. The kind of domain-specific knowledge we specified as core theory. In view of our main constructs, we think three aspects of theoretical knowledge are important in particular. First, there is knowledge of socio-historical developments of a construct in relation to the teaching and learning of other languages. Second, there is familiarity with key interpretations and specifications of the construct and of some important related notions. Finally, there is knowledge of recent research interests and findings.

Having said this, we will not ever forget how difficult it is for teachers to find the time and opportunities to keep up with domain-specific knowledge, given the time and energy their profession calls for throughout a school year. Spend a school year with Joy, Mark and Pete, and anyone will understand what we mean. We again stress that unless favourable conditions are created, the majority of teachers will not be able to keep their subject-knowledge up to date and critically relate what they have learned to their teaching and testing practices.

Yet, when adolescent learners are taught how to communicate in English and develop their autonomy at the same time, the areas and issues we have so far discussed should preferably not be addressed in isolation of one another. Instead, we would like to suggest explorations of notions that relate to all of the three main constructs of this investigation.

The next section will discuss two research paradigms that typically address the investigation of complex social practices in real-world contexts.

12.4.4 Involving teachers in reflective action and design-based research

The fourth and final of our recommendations is to have teachers systematically investigate their own teaching practices in cooperation with others. Following Lewin (1948), we wish teachers to be researchers. The type of research we have in mind requires paradigms suitable for exploring and investigating the ecology of classroom teaching and learning. We feel two paradigms are particularly relevant in this respect: *reflective action research* and educational *design research*, more frequently called *design-based research*. The paradigms hold the promise to help teachers as well as researchers relate practice to theory and theory to practice. The types of research may lead to practitioners and theorists in close cooperation becoming agents of educational innovation and change. We will briefly discuss the two paradigms below.

Reflective action research

Reflective action research involves systematic investigations initiated by teachers who wish to improve their teaching practices by understanding them more fully, in collaboration with others. (Curry, 2006; Flamini & Jiménez-Raya, 2007). Carrying out reflective action research is meant to lead to teachers' professional development and observable improvements of their teaching and testing practices. Action research is generally seen to consist of six related steps.

The starting point is identifying an issue, problem, or situation a teacher is engaged to investigate by way of questions such as 'What happens when I ….?' This is already a stage where academic theory and recent research results may play a role. Curry (2006) offers specific examples of questions that are asked at the beginning of an action research cycle: 'What happens when I use dialog journals in a writing class?' or 'What happens when I use authentic materials, such as weblogs, websites, TV shows, and magazines, to ground instruction in popular culture?'.

The second step is an attempt to clarify, narrow or focus the initial question. Curry (2006) mentions the example of Alexandra, one of her pre-service student teachers, who was broadly interested in questions on how to respond to her learners' writing. After reviewing some of the literature on feedback to writing, Alexandra came to specify her initial question to 'What happens when I respond in three different ways (direct correction of errors, circling errors without correcting them, and providing holistic feedback) to student writing?'.

The third step in action research is to define the data collection context, timeframes and methods. Alexandra, thus, needed to specify beforehand which writing assignments offered to the learners would receive which of the three types of feedback, how she intended to compare her learners' responses to her feedback, and how she would elicit her learners' views on the different types of feedback. In the context of Alexandra's question, data collection might have included scores on previous tests, diaries or journals kept by her and/or her learners, talk-aloud protocols, observations, recordings, questionnaires, interviews, and even quasi-experiments.

In the fourth step, the teacher analyses the data and looks for changes from previous behaviours or practices, or identifies recurrent patterns or themes. In this stage, preferably several sources of data will be triangulated and their results subsequently interpreted by the teacher.

Fifth, action strategies are developed and put into practice at the first given opportunity, at which point the effectiveness of the new strategies can be investigated using the same research cycle. Here, theory may again be helpful to refine, restructure or further specify the action strategies.

The final step of reflective action research is dissemination, discussion and dialogue. The new knowledge, skills, understandings or action strategies should be made public and presented to and discussed with peers, school colleagues, in professional development workshops, in newsletters or at conferences. Discussions

and evaluations are needed to further validate the action research findings and arrive at new perspectives for research cycles to come.

Flamini & Jiménez-Raya highlight the importance of action research for a teacher's professional development and deduce the following characteristics from the various definitions an interpretations of action research they had found. Action research is *critical, practical, small-scale, collaborative, participatory, reflective and dialectical, cyclic, democratic and equitable,* and finally, *experiential* (2007:106/9).

Design-based research

Particularly because this study has shown how little informed the teaching and testing practices of our respondents were by academic theory, a research paradigm different from reflective action research may be adopted when teachers critically investigate their communicative teaching and testing practices.

Design-based research, also referred to as (educational) design research, typically attempts to blend academic theories, practical theories, and curriculum generation in projects in which theorists and practitioners closely cooperate. The essence of design-based research is that *research* and *design* are mutually developed by researchers and practitioners cooperating in real-world settings. According to Cobbe et al. (2003:10), a design experiment in educational research draws on prior research and attempts to cash in on the empirical and theoretical results of that research. Projects are planned and carried out with the aim to develop a class of theories both about the process of learning and of the means designed to support the targeted learning. Examples of such supports of learning are particular forms of learner behaviour, learning strategies or didactic procedures, which are developed and repeatedly put to the test in the curricular implementation of disciplinary ideas about domain-specific constructs, such as communicative ability and its formative assessment and evaluation.

Wang and Hannafin (2005) highlight five characteristics of design-based research in the literature. They first refer to design-based research (DBR) as being *pragmatic*, because it focuses on solving real-world problems by designing and enacting interventions as well as extending prior theories and redefining design principles (Design-based research collective, 2003; Van den Akker, Gravemeijer, McKenney, & Nieveen (2006) . Its pragmatic quality is also shown by the participants' efforts to disseminate and further discuss and develop their findings.

Secondly, DBR is firmly *grounded* in both theory and a real-world context, amidst the dynamics, complexities, opportunities and limitations of authentic practice. It is not unlike the efforts of the Wright brothers, who actually designed and improved planes while flying them, and by doing so added to the very concept of what a plane is and what it can be used for.

Thirdly, DBR is essentially *interactive, flexible* and *iterative*. Teachers and researchers interact, bringing with them their specific expertise. Together, they flexibly design and evaluate repeated cycles of the research theme selected, until these very cycles tend to become repeatable.

Fourthly, DBR is *integrative*, because it requires that participants integrate a variety of research methods and approaches form both qualitative and quantitative research paradigms, depending on the needs of the investigation. In addition to its multiple sources of analysis, this makes DBR methodologically challenging. Van den Akker, Gravemeijer, McKenney, & Nieveen (2006) have provided an introduction to the issues of design research as a methodology, in addition to offering and discussing examples from the field.

Finally, DBR is *contextual*, with the setting where the research is conducted constantly interacting with the design process through which results are generated (Wang & Hannifin, 2005:11). It requires extensive documentation and detailed records of the design research process.

We feel the findings of the present study has provided relevant issues, themes and notions as input for the research paradigms we have discussed. The issues, themes and critical concepts we raised regarding the promotion of LA, CLE and effective assessment and evaluation are waiting to be explored and developed any further.

We sincerely hope that our four recommendations are taken up by anyone with a vested interest in the promotion of learning in formal educational settings and in the means with which such learning can be facilitated.

Even though our interest primarily was in self-regulated learning when adolescents learn how to communicate in another language, we feel some of our notions and recommendations go beyond the domain-specificity of the subject of English as a foreign language. This has particularly been the case for notions related to LA and assessment and evaluation.

In the final section of our discussion chapter we will go into eleven shortcomings of the investigation that lead to modesty and care in the interpretation of our findings.

12.5 Shortcomings

Looking back on our quest for autonomy, we feel that the following issues require attention and call for improvement in future investigations.

Initial research plan and objectives too widely formulated and too ambitious

When we started our investigations by writing a fundable research plan, we did so with more enthusiasm than knowledge. Next to practical experiences as a teacher and teacher educator, our knowledge did not amount to much more than general and domain-specific knowledge of the three constructs we were challenged to investigate in a school setting. In addition, our knowledge of research methodology was either outdated, or was to be built up from the start.

Lack of knowledge of exploratory multiple-case designs led to a research plan with objectives that appeared to be too widely formulated and too ambitious. That is why the initial research questions soon had to be specified in no less than seven related questions. It was not only difficult for Joy, Mark and Pete to specify what they were after. It was also difficult for the researcher to specify his non-participatory, exploratory observations and reflections.

We nevertheless feel that the knowledge and experiences we gathered in the course of the study, will be helpful in making future research projects perhaps less ambitious, but certainly more specified and focused.

Data- analyses too detailed and time-consuming at first

In our first part-time data analyses in-between teaching, we rather rigorously stuck to the principles of grounded theory. This meant that we painstakingly analysed the teacher interviews in an attempt to capture every possible clue and every nuance that helped us reflect on our expanding research questions. It was almost as if the semi-structure of our interview guides was not taken into account in our first series of data analyses. Of course, looking for every possible clue and nuance is and remains a necessary step in grounded theory. However, we appeared to overshoot the mark, so that it took us too long to condense the data to the general patterns identified in chapters 11 and 12. Besides, there were some persistent technical problems with the software we used when analysing our data, which did not exactly speed up the process of analysis.

The credibility of non-convergence of data

Non-convergence of a phenomenon does not with positivist certainty mean that some crucial event did not occur at some time in the year of data collection.

Researchers have to be strict about the internal and external validity and the reliability of their findings. We argued in chapter 2 that validity and reliability in a strictly positivist sense can never be achieved in an exploratory multiple-case study, such as the one we carried out. Instead, we did all we possibly could to go for the credibility of patterns in the respondents' perceptions, as they persistently recurred in the course of the school year. Nevertheless, we feel that more detailed and more specific evidence could have converged in our study.

As a part-time researcher, we did not have the opportunity to follow our respondents on a week-to-week basis. This means that relevant episodes or events may have occurred in the classrooms we had no knowledge of, because they were neither explicitly reported on by the informants in the interviews we had, nor came to the fore in our limited classroom observations. Even though the actual teacher-made tests successfully managed to elicit the teachers' beliefs, in retrospect we would have liked to have collected more detailed data on the ways in which the respondents interpreted the constructs, by way of more opportunities to:

- be present at the schools;
- check, analyse and reflect on the data in-between school visits;
- reflect on the data with fellow researchers in the course of data collection by way of parallel analyses, discussion and negotiation;
- videotape classroom observations with a clear research focus;
- gather teacher as well as learner data before, during and immediately after the tests the respondents had selected.

The respondents' representativeness

We realise that our teacher respondents cannot be called representative of their colleagues in the Netherlands, or elsewhere in the world for that matter. Representativeness of our respondents in a positivist sense had never been on our mind. Besides, all of the respondents were in some way or other connected with ILS, the graduate school of education of Radboud University, at which the present researcher was employed.

Our aim had always been to select interesting teachers without class management problems who were expected to have developed effective teaching and testing practices. We selected our respondents in the hope that they would provide a wide variety of teacher data. In view of this variety, it is regrettable that Joy and Pete were not only employed at the same school, but were also used to cooperating closely. Pete's practical theories, projects and testing practices were admired and endorsed by Joy. As a result, the two respondents saw eye to eye in many respects, so that patterns between these two respondents were more likely to occur. Moreover, Pete consistently accepted Joy's choice for a test that was offered for discussion. This means that of the original 9 tests we envisaged, only six came to be the focus of our investigation. On a more positive note, we might add that it was interesting to compare two colleagues working together and separately interviewing them on the same tests.

We still maintain that the teachers we had selected were good-practice teachers. We have come to experience that Joy, Mark and Pete are educators one can safely send one's children to, without any serious worries about the level of English they offer their learners. We feel that more representative teachers, and probably more average teachers, than Joy, Mark and Pete would likely to have had class management problems. Such problems might have seriously interfered with the research results. Similarly, more representative teachers would probably not have constructed as many tests of their own as our three respondents did in the course of the year. After all, our main source of evidence was to consist of teacher-made tests, and not of tests that would come with the course materials and would be used by the teachers without much conscious thought. What we suggest here is that other

teaching and testing practices might have produced less positive data on teaching and testing than was the case with our informants.

Nevertheless, the selection of participants will call for more serious attention in research projects to come.

Absence of formal learner data

Although learner data were originally envisaged in the research plan of this study, it soon appeared to be too ambitious to also include the data of learner questionnaires and focus-group interviews in the study. We nevertheless feel that detailed learner data would have added to both the quality and credibility of our findings, and we will most certainly include these in projects to come.

A focus on CLE should have included oral as well as written tests

From the start, our focus had been on written tests. This implied that the tests the respondents were to select for discussion were likely to be tests of writing ability. In order to understand our respondents' interpretation of communicative competence more fully, it would have been better to also have focused on the ways in which the respondents taught and tested oral skills.

The dilemma of credibly as well as concisely reporting on qualitative studies

One of the problems that particularly qualitative researchers have to face is how to report on their investigations in credible ways, and at the same time be concise and to the point. If qualitative researchers wish their investigations to be credible, they need to provide convincing evidence of what they have done and on the basis of which data conclusions or recommendations have been made.

Unfortunately, we needed more pages and chapters than originally envisaged to arrive at an acceptable level of credibility.

Convincingly and concisely reporting on qualitative studies will remain a challenge in times to come.

APPENDIX I: HAVO CURRICULUM

The standard study load for the 4th and 5th years of HAVO combined amounts to:

- 1,480 hours for the common component;
- 1,160 hours for the specialised component;
- 560 hours for the optional component.

These hours are based on the time required by the average pupil. Unlike the recommended timetable for basic secondary education basisvorming where each "hour" equates with a teaching period of 50 minutes, these hours are "real" hours, i.e. 60 minutes. The study load per subject is shown below.

HAVO subject combinations

The common component, specialised components and optional component comprise the following subjects:

Common component (for all pupils)

Dutch language and literature	400
English language and literature	360
French, German, Spanish, Russian, Italian	160
Arabic, Turkish or Frisian language and literature 1	160
General science	160
history and social studies	120
culture and the arts	120
physical education	

Science and technology (specialised component)

mathematics B1&2	440
physics 1&2	440
chemistry	280

Science and health (specialised component)

mathematics B1	320
physics 1	240
chemistry	280
biology	320

Pupils may, if they wish, exchange the half-subjects mathematics B1 and physics 1 for mathematics B1&2 and physics 1&2 respectively, both of which are full subjects.

Economics and society (specialised component)

economics 1&2	440
mathematics A1&2	280
history	240
geography	200

Pupils may, if they wish, swap the full subject mathematics A1&2 for the half-subject mathematics B1 or the full subject mathematics B1&2.

Culture and society (specialised component)

- one of the following half-subjects, to be combined with the corresponding half-subject from the common component:
  - French, German, Spanish, Russian, Italian, Arabic, Turkish or Frisian language or literature 2: 200

or one of the following subjects:

  - French, German, Spanish, Russian, Italian, Arabic, Turkish or Frisian language and literature 1&2: 360
- the half-subject culture and the arts 2 with a study load of 120 hours, combined with one of the following half-subjects with a study load as indicated:
  - culture and the arts 3 (art and design) 240
  - culture and the arts 3 (music) 240
  - culture and the arts 3 (drama) 240
  - culture and the arts 3 (dance) 240

- history 240
- economics 1 200
- mathematics A1 160

Pupils may, if they wish, swap the half-subject economics 1 for the full subject economics 1&2 and swap the half-subject mathematics A1 for mathematics A1&2 or mathematics B1&2 (both full subjects) or the half-subject mathematics B1.

Depending on the choices made, the total study load for pupils opting for science and health, economics and society or culture and society may be greater than the figure of 1,160 hours mentioned above.

Optional component

- any of the subjects or half-subjects listed for the specialised components; if the half-subject mathematics A1 or the full subject mathematics A1&2 is combined with the half-subject mathematics B1 or the full subject mathematics B1&2, the study load for the subject or half-subject mathematics, chosen as part of the optional component, is reduced by 160 hours.
- any of the following subjects (for pupils who did not study the language concerned during the period of basic secondary education):

- Spanish language and literature (for beginners) 360
- Russian language and literature (for beginners) 360
- Italian language and literature (for beginners) 360
- Arabic language and literature (for beginners) 360
- Turkish language and literature (for beginners) 360

- any of the following subjects:
  - social studies 200
  - philosophy 360
  - management and organisation 280
  - information technology 240
  - physical education 2 240

- subjects or other elements of the curriculum specified by the competent authority
- any of the following half-subjects:
  - French language and literature 1 160
  - German language and literature 1 160
  - Spanish language and literature 1 160
  - Russian language and literature 1 160
  - Italian language and literature 1 160
  - Arabic language and literature 1 160
  - Turkish language and literature 1 160
  - Frisian language and literature 1 160

- subjects and half-subjects listed under the specialised components for VWO which do not correspond or overlap with subjects or half-subjects for HAVO.

Recommended number of teaching periods for HAVO (old-style)

The old and new systems will continue to exist alongside each other until the end of the 2002/2003 school year. The old-style HAVO examination, based on the recommended number of teaching periods for HAVO, will be sat for the last time in 2003.

Future developments

The HAVO curriculum has been found to be overloaded and fragmented. An adapted curriculum will therefore be introduced as from August 2006. The following changes have been proposed within the existing structure:

- half-subjects will become full subjects;
- schools and pupils will have more freedom of choice;
- there will be far fewer regulations relating to examinations, especially as regards the school exam.

APPENDIX II: VWO CURRICULUM

The standard study load for the 4th, 5th and 6th years of VWO combined amounts to:

- 1,960 hours for the common component;
- 1,840 hours for the specialised component;
- 1,000 hours for the optional component.

These hours are based on the time required by the average pupil. Unlike the recommended timetable for basic secondary education basisvorming where each "hour" equates with a teaching period of 50 minutes, these hours are "real" hours, i.e. 60 minutes. The study load per subject is shown below.
VWO subject combinations
The common component, specialised components and optional component comprise the following subjects and half-subjects:

Common component (for all pupils)

- Dutch language and literature 480
- English language and literature 400
- French language and literature 1 160
- German language and literature 1 160
- general science 200
- history and social studies 200
- culture and the arts 1 200
- physical education 1 160

Science and technology (specialised component)

- mathematics B1&2 760
- physics 1&2 560
- chemistry 1&2 520

Science and health (specialised component)

- mathematics B1 600
- physics 1 360
- chemistry 1 400
- biology 1&2 480

Pupils may, if they wish, swap the half-subjects mathematics B1, physics 1 and chemistry 1 for mathematics B1&2, physics 1&2 and chemistry 1&2 respectively, all of which are full subjects.
Economics and society (specialised component)

- economics 1&2 520
- mathematics A1&2 600
- history 360
- geography 360

Pupils may, if they wish, swap the full subject mathematics A1&2 for the half-subject mathematics B1 or the full subject mathematics B1&2.
Culture and society (specialised component)

- one of the following half-subjects, to be combined with the corresponding half-subject from the common component:
    - French language and literature 2 200
    - German language and literature 2 200

or one of the following subjects:

    - Spanish language and literature 480
    - Russian language and literature 480
    - Italian language and literature 480
    - Arabic language and literature 480
    - Turkish language and literature 480
    - Frisian language and literature 400
    - Latin language and literature 480
    - Greek language and literature 480

- one of the subjects or half-subjects listed under a) with the study load indicated, or philosophy with a study load of 320 hours

- the half-subject culture and the arts 2 with a study load of 200 hours, combined with one of the following half-subjects with a study load as indicated: 200
    - culture and the arts 3 (art and design) 280
    - culture and the arts 3 (music) 280
    - culture and the arts 3 (drama) 280
    - culture and the arts 3 (dance) 280

- history 360
- mathematics A1 360

Pupils may, if they wish, swap the half-subject mathematics A1 for mathematics A1&2 or mathematics B1&2 (both full subjects) or the half-subject mathematics B1.

Depending on the choices made, the total study load for pupils opting for science and health, economics and society or culture and society may be greater than the figure of 1,840 hours mentioned above.

Optional component

The optional component for all subject combinations at VWO level may comprise:

- any of the subjects or half-subjects listed for the specialised components; if the half-subject mathematics A1 or the full subject mathematics A1&2 is combined with the half-subject mathematics B1 or the full subject mathematics B1&2, the study load for the subject or half-subject mathematics, chosen as part of the optional component, is reduced by 280 hours.
- any of the following subjects (for pupils who did not study the language concerned during the period of basic secondary education):
    - Spanish language and literature (for beginners) 480
    - Russian language and literature (for beginners) 480
    - Italian language and literature (for beginners) 480
    - Arabic language and literature (for beginners) 480
    - Turkish language and literature (for beginners) 480
- any of the following half-subjects:
    - biology 1 160
    - economics 1 280

- any of the following subjects:
  - social studies 360
  - management and organisation 360
  - information technology 280
  - physical education 2 280

- classical culture with a study load of 200 hours (for pupils studying Latin language and literature and/or Greek language and literature)
- subjects or other elements of the curriculum specified by the competent authority.

All gymnasium pupils study classical culture instead of the half-subject culture and the arts 1, whatever the subject combination chosen. They may, however, opt for culture and the arts 1 as part of the optional component. All "gymnasium" pupils also take Latin language and literature and/or Greek language and literature, both of which have a study load of 480 hours.

Recommended number of teaching periods for VWO

The old and new systems will continue to exist alongside each other until the end of the 2003/2004 school year. The old-style VWO examination will be sat for the last time in 2004.

Future developments

The curriculum has been found to be overloaded and fragmented. An adapted curriculum will therefore be introduced as from August 2006. The following changes have been proposed within the existing structure:

- half-subjects will become full subjects;
- schools and pupils will have more freedom of choice;
- there will be far fewer regulations relating to examinations, especially as regards the school exam.

APPENDIX III: EXAMINATION PROGRAMME ENGLISH LANGUAGE AND LITERATURE VWO

1 The final examination

The final examination consists of the national examination and the school examination.

The examination programme comprises the following domains:

- Domain A Reading skills
- Domain B Listening skills
- Domain C Speaking skills
- Domain D Writing skills
- Domain E Literature
- Domain F Orientation on further studies and careers

1.1 The national examination

The national examination relates to reading skills (domain A), subdomain ‘language skills’.

The national reading examination is administered in a 2,5-hour session.

1.2 The school examination

The school examination consists of an examination file with evidence of the levels of achievement of the attainment targets. The file comprises three components: a selection of tests and their results, an activities component, and a profile assignment.

Part a Tests

Listening skills

Listening is tested by way of a varied selection of texts with questions and tasks that examine the attainment targets of the subdomain of language skills.

Speaking skills

Speaking is tested by way of a varied selection of texts with questions and tasks that examine the attainment targets of the subdomain of language skills.

Writing skills

Writing is tested by way of a varied selection of texts with questions and tasks that examine the attainment targets of the subdomain of language skills.

Literature

The subdomains 'literary notions' and 'literary history' are either assessed orally or in writing by way of texts. Integrative testing of literature and language skills is not allowed. The departments of *Dutch language and literature*, *modern languages and literature*, and *cultural and arts education 1* are expected to have aligned assessment procedures.

Assessment is based on scoring criteria, keys that present appropriate response, or an assessment model with scoring criteria. Speaking and writing are assessed by way of the aspects mentioned under ‘Assessment’. Candidates are familiar with the assessment criteria before the actual assessment takes place.

Part b Activities component

Reading skills

The candidate has done tasks that involved extensive reading, summarising texts, and gathering information with the help of information and communication technology.

Listening skills

The candidate has done tasks that involved extensive listening and note-taking .

Speaking skills

The candidate has used the target language in realistic communicative situations a number of times and has done some presentations in the target language.

Writing skills

The candidate has used the target language in written correspondence, also with the help of telecommunication, and has experience in writing reports.

Literature

The candidate has compiled a reading file, in which experiences with literary texts and essays have been documented.

The reading file at least includes the tasks mentioned under attainment target 43 and a list of references to the literature or other sources consulted. Variation in tasks is required. In addition, if a school so wishes, the reading file may include a reading autobiography and/or one or more evaluative reading reports.

- In a reading autobiography, the candidate describes his development as a reader up to the moment of writing.
- In an evaluative reading report, the candidate evaluates his/her reading- and learning process over time.

Orientation on further studies and careers

The candidate has gathered information on further education and professions that require foreign languages or demand some specific use of them.

No scores or grades will be assigned to the tasks of the activities component. All of the activities must simply be passed.

Part c Profile assignment

(only for candidates with a culture and society profile)

The profile assignment requires a study load of 80 hours. It relates to at least two (partial)subjects of the profile curriculum (which includes Dutch and English language and literature).

If English language and literature is involved, the profile paper entails:

- an independent research design
- the investigation of a linguistic or literary concern.

The requirements of the profile assignment are not met if the candidate has limited him/herself to the use of English alone.

The work is documented or presented in one of the following ways:
- a written report (research paper, review, narrative report, report on a questionnaire or interview);
- an essay or article (exposition, contemplation, or argumentation);
- an oral presentation (exposition, contemplation, argumentation, or panel discussion);
- a set of propositions with argumentation;
- an illustrated poster presentation;
- a presentation in which media are used (audio, video, ICT).

The paper or oral presentation includes reflections on the learning process (rationale, research questions, activities carried out, sources consulted, etc.). These reflections will be assessed.

Assessment is done on the basis of scoring criteria, which have been made known to the candidate beforehand. The papers and/or presentations are scored by the examiners of the profile subjects on which the assignment has been based.

Acceptable papers or presentations are graded either ‘sufficient’ or ‘good’.

Information and communication technology (ICT)

The candidate is invited to make use of the following ICT applications:
- consultation of (hyper)texts, data, images and sound in (multimedial) data, data banks and information systems retrieved with the help of a computer (network);
- search engines of libraries and media centres;
- telecommunication, such as e-mail, discussion and news groups;
- word processing;
- electronic or graphic calculator;
- mathematical operations;
- spreadsheets, models and simulations;
- retrieval and management of data from data banks and information systems;
- design of (multimedial) presentations;

The use of ICT-application in testing is optional and dependent on the hardware and/or software available at a school.

Weighting

Part a:

The scores on listening, speaking, and writing skills together determine one third of the final school examination mark.

Literature is scored separately, in combination with the mark received for Dutch literature. The mark for Dutch literature has a weighting factor of 3 and the mark for literature for each of the foreign languages (subdomains ‘ literary notions’ and ‘literary history’) has a weighting factor of 1.

Literary development of modern foreign languages is part of the activities component.

Part b:

The actitivies component does not contribute to the school examination mark.

Part c:

The scores on the profile assignment result in a separate mark on the examination list.

2. Examination contents

The examination contents have been described in terms of attainment targets and related to the 5 levels that have been specified in appendix 19.

2.1 Attainment targets

Domain A: Reading skills
Subdomain: Language skills

The candidate is able to:
1 indicate whether, given a certain need, a text contains relevant information, and if so, what or which information.
2 point out the main idea of a text or part of a text.
3 indicate the meaning of important text elements.
4 label parts of a text and indicate logical connections between them.
5 draw conclusions on the basis of the text, in view of the target audience, language use, objectives, and the author's views and feelings as well.

Attainment targets 1-5 concern texts that are read in view of social contacts, personal needs, and study. The candidate is expected to adjust his/her reading strategies to the respective reading targets.

Subdomain: Extensive reading

6 The candidate has gained wide experience in reading texts that are easily readable.

Subdomain: General skills

The candidate has:
7 gained experience of the technique of summarising as a way to tackle a text.
8 used ICT applications in acquiring information several times.

Attainment targets 1-8 are assessed on the basis of level 4 specifications of reading skills.

Domain B: Listening skills
Subdomain: Language skills

The candidate is able to:

9 indicate whether, given a certain need, a spoken text contains relevant information, and if so, what or which information.
10 indicate the meaning of important text elements.
11 label parts of a text and indicate logical connections between them.
12 draw conclusions on the basis of the text, in view of the speaker's intentions, views and feelings.
13 anticipate the most likely continuation of a conversation on the basis of what one has heard.

Attainment targets 9-13 concern spoken texts that are listened to in view of social contacts, personal needs, and study. The candidate is expected to adjust his/her listening strategies to the respective listening targets.

Subdomain: Extensive listening

14 The candidate has gained wide experience in listening to recorded sound and/or visual materials.

Subdomain: General skills

15 The candidate has gained experience of the technique of note-taking as a strategy to tackle spoken texts.

Attainment targets 9-15 are assessed on the basis of level 4 specifications of listening skills.

Domain C: Speaking skills
Subdomain: Language skills

The candidate is able to:

16 start and finish a monologue or conversation.
17 call for attention and ask for clarification or repetition.
18 thank, apologise, congratulate, and invite.
19 give and ask for information.
20 give an opinion and ask for one.
21 describe someone or something.
22 express and ask for feelings, interests, and preferences.
23 comment and pass judgement.
24 plead, complain, and negotiate.
25 instruct.
26 to express and justify a point of view.

Attainment targets 16-26 are assessed on the basis of level 4 specifications of speaking skills.

Subdomain: Practical situations

27 The candidate has used the target language a number of times in practical situations.

Subdomain: General skills

28 The candidate has given presentations in the target language a number of times.

Assessment

The following aspects need to be considered: content (completeness, detail, comprehension), correct and appropriate language use, fluency, and pronunciation.

Domain D: Writing skills
Subdomain: Language skills

The candidate is able to:
29 thank, apologise, congratulate, and invite.
30 give and ask for information.
31 give an opinion and ask for one.
32 describe someone or something.
33 express and ask for feelings, interests, and preferences.
34 comment and pass judgement.
35 plead and complain.
36 to express and justify a point of view

Attainment targets 29-36 are assessed on the basis of level 4 specifications of writing skills.

Subdomain: Practical situations

37 The candidate has participated in correspondence projects, if needed by way of telecommunication.

Subdomain: General skills

38 The candidate is able to use the facilities offered by word processing when writing.
39 The candidate has written a written a report in the target language, for the benefit of a project or another school subject.

Assessment

The following aspects need to be considered: content (completeness, originality, clarity), structure, and correct and appropriate language use.

Domain E: Literature
Subdomain: Literary development

40 The candidate manages to learn from a varied selection of texts, so that (s)he is able to develop a taste for reading that connects with his/her personal preferences.
41 By way of literary texts, the candidate manages to learn about a number of aspects of society, which help to develop interpretations reality and stimulate reflections on his/her position in the real world.
42 On the basis of the learning experiences referred to in targets 40 and 41 above, the candidate is able to report on and justify a personal selection of three of the literary works (s)he has read in the target language.

- The literary works are original works.
- The texts are within the candidate's grasp, which does not mean that the texts are exclusively selected from juvenile literature; however, the works should be accessible to the candidates.
- In some cases, translations can be used next to the original versions.

43 The candidate is able to describe, reflect on, and evaluate the personal learning experiences that derive from a number of (sections of) literary works.

- The *description*:
  includes the candidate's personal response to the works, justification of choice, and brief summary of the contents.
- The *reflection* is related to a specific task and may involve a(n):
  - discussion of crucial extracts;
  - discussion of activities related to reading, such as the ways the reader's expectations have been built up or the ways in which the reader has been invited to identify with certain characters;
  - personal reflection on the candidate's response to the text or to the background information made available;
  - comparison of the candidate's learning experiences with his/her peers or readers allegedly with more expertise (critics, teachers);
  - discussion of the characters;
  - analysis of the ways in which tension has been built up;
  - discussion of the work in relation to the author's biography or beliefs;
  - comparison of the work with other works written by the author;
  - comparison with other authors or different literary works;
  - discussion in relation to a particular cultural-historical or social context;
- The evaluation includes a final assessment of the work and an evaluation of the candidate's personal learning experiences and reflections, in which among other things (s)he indicates what was difficult, confusing, or vague.

Remark: The description, reflection, and evaluation of (extracts from) literary works can be based on a common theme or social aspect.

Subdomain: Literary notions

The candidate is able to:

44 distinguish literary test types and indicate in what ways they differ from non-literary and non-fictional types of text.

45 make use of the knowledge base of literary notions that are mentioned in the examination programme for Dutch language and literature.

Remark: It should be stressed that the knowledge base of literary notions is intended to enhance reflection and discussion of literary works. Therefore, it cannot be examined separately.

46 analyse, interpret and assess (extracts from) literary works with the help of the literary notions referred to in attainment target 45.

Subdomain: Literary history

The candidate:

47
- has some knowledge of narrative prose, dramatic art, and poetry from a sustained and relatively long literary-historical period, in addition to knowledge of contemporary (prose)literature.
- masters the knowledge required to interpret longer literary-historical periods, such as knowledge of important literary movements, authors, their lives, their works, and their relevance to the development of contemporary culture. In addition, the candidate is attentive to international connections.

48 is able to recognise and mention characteristics of movements, authors and (sub)genres that relate to the period mentioned in attainment target 47, on the basis of his/her knowledge of narrative prose, dramatic art, and poetry of that period.

Domain F Orientation on further studies and careers

49 The candidate has gathered information on types of further education that require foreign languages or demand some specific use of them.

50 The candidate has determined in how far (s)he has the attitude, interest, and skills that are required for further education.

APPENDIX IV: INTERVIEW GUIDE USED IN THE FIRST INTERVIEWS

<u>Interview-guide</u> t.b.v het vrije-attitude interview met eerstegraads docenten Engels over hun toetsingspraktijk gericht op (deel)-aspecten van taalvaardigheid in de Engelse taal.

Introductie

Vooraf
Toetsen kunnen het leren beïnvloeden. De toetsen worden voorbereid, gemaakt en veelal geëvalueerd door de verantwoordelijke docenten en door hun leerlingen. Met name de toetsen waarvoor een cijfer of andere formele beoordeling wordt gegeven, spelen een rol in het leerproces van de leerling in het voortgezet onderwijs. Dit onderwijs is in toenemende mate gericht op het bevorderen van de zelfstandigheid van de leerling. De rol die toetsing speelt in het leerproces is complex. Toetsen kunnen zowel een positief als een negatief effect hebben op het leren van de leerling. Om geldige uitspraken te doen over de rol van toetsing in het leerproces is meer kennis vereist over ***wat*** docenten Engels aan taalvaardigheid toetsen, ***hoe*** zij dit doen en ***waarom*** zij dit zo doen.

Ik heb u benaderd, omdat leerlingen, ouders, c.q. verzorgers en schoolleiding u waarderen als docent Engels. Tevens bent u actief betrokken bij de invoering van de vernieuwde Tweede Fase.

Ter introductie ga ik eerst iets vertellen over het doel, de werkwijze en het vervolg van het eerste vraaggesprek dat wij samen voeren. Daarna beginnen we met het vraaggesprek zelf.

Doel
Doel van het interview is informatie te verkrijgen over uw visie op toetsing in uw dagelijkse onderwijspraktijk. Daarbij komen tenminste de volgende vijf aspecten aan bod. In de eerste plaats zullen we spreken over de **[a] ontwikkelingen** in uw toetsingspraktijk vanaf het moment dat u voor het eerst met het fenomeen toetsing werd geconfronteerd tot aan de stand van zaken op dit moment. Tevens zal ik informeren naar **[b]** wat u verstaat onder een **goede schriftelijke taaltoets**. We zullen dan doorpraten over de **kennis** van, het **inzicht** in en de **vaardigheden** in de Engelse taal die u voor uw leerlingen van belang acht. Ten derde **[c]** vraag ik u aan te geven wat u verstaat onder drie kernbegrippen die een rol spelen in het huidige moderne vreemde talen onderwijs: **communicatief taalonderwijs**, **zelfstandig leren** en **zelfverantwoordelijk leren**. Ten vierde vraag ik u naar **[d]** de rol die uw toetsvragen en **toetsen** spelen in het leren communiceren van de individuele leerling in het Engels. We sluiten af met de vraag naar **[e] besluiten en voornemens** voor uw toetsingspraktijk in het schooljaar 1999-2000 in het vierde leerjaar van het havo en vwo.

Tijd en werkwijze
Het gesprek duurt circa 90 minuten. Ik wil benadrukken dat het gesprek is gericht op weergave van uw persoonlijke mening en ervaringen. Verkeerde antwoorden van uw kant zijn dus ***niet*** mogelijk Er vindt ***geen*** discussie plaats. De vragen die ik binnen de vijf kaders stel, bewegen zich van algemeen naar meer specifiek en eindigen regelmatig met door u gegeven voorbeelden van wat u bedoelt.

Voordat we met het bespreken van een aspect beginnen, vraag ik u om in maximaal

vijf minuten in sleutelwoorden op te schrijven wat u binnen de vijf kaders zelf aan de orde wilt stellen. Het is belangrijk dat al deze sleutelwoorden in de loop van het gesprek over het betreffende aspect aan bod komen.

Na afloop van het gesprek probeer ik het verloop samen te vatten en spreken we af dat u voorbeelden van toetsvragen en/of toetsen levert ter illustratie van de besproken onderwerpen. Zoals is afgesproken kunt u gegevens uit uw docentenagenda van nu en/of van afgelopen jaren gebruiken als geheugensteun of ter illustratie.

Opname
Met het oog op de verwerking van de gegevens van dit interview is het noodzakelijk dat ik een audio-opname van het vraaggesprek maak. De gegevens zullen anoniem worden verwerkt., d.w.z zodanig dat u niet herkenbaar zult zijn. Uw naam wordt vervangen door een code en eventuele andere namen die u noemt worden vervangen door een hoofdletter. Verder wil ik benadrukken dat de gegevens uit dit interview strikt vertrouwelijk behandeld zullen worden. De bandopnames en de uitgeschreven teksten zullen uitsluitend binnen het onderzoekskader worden beluisterd of gelezen en over hetgeen u vertelt wordt niets aan derden medegedeeld.

Verwerking
De gegevens van dit vraaggesprek en de door u ter illustratie geleverde toetsvragen en/of toetsen worden door mij beschreven en geanalyseerd. De uitkomsten van de beschrijvingen en analyses worden in een tweede vraaggesprek aan u voorgelegd. (***wishful thinking***)

Afsluiting inleiding
Als u vragen hebt, stel deze dan gerust. Vervolgens vraag ik u in maximaal 5 minuten in sleutelwoorden op te schrijven wat u ter sprake wilt brengen **binnen de vijf deelgebieden** van het interview.

APPENDIX V: INTERVIEW GUIDE USED IN THE FOLLOW-UP INTERVIEWS

Introductie

Vooraf

Dit is het tweede/derde vraaggesprek dat we samen voeren. Ik neem het gesprek weer op. We gaan spreken over de derde toets die je hebt uitgekozen. De toets is door jou getypeerd als een toets die inzicht geeft in de kennis van en vaardigheden in het Engels van de leerlingen uit jouw vierde klas. Je hebt gekozen voor ……..

Doel

Doel van het interview is informatie te verkrijgen over de geselecteerde toets. Daarbij komen in ieder geval de volgende vijf aspecten aan bod. Op de eerste plaats zullen we spreken over de **[a] motivering** om juist voor deze toets te kiezen. Vervolgens gaan we praten over de **[b] de kennis** en de **vaardigheden** die en **het inzicht** dat je met de toets meet. Dan volgen meer specifieke vragen over **[c] constructie en gebruik** van de toets, zoals voorbereiding op de afname, afnamecondities, de interpretatie van de behaalde scores en de eventuele nabespreking met de klas of met individuele leerlingen. Dan vraag ik in hoeverre de constructie en het gebruik van de toets het **[d] luisteren, spreken, lezen en schrijven in het Engels** bevordert. We sluiten af met de vraag in hoeverre constructie en gebruik van de toets aanzet tot **[e] zelfstandig en/of zelfverantwoordelijk leren.**

Tijd en werkwijze

Het gesprek neemt circa 90 minuten in beslag. Net als bij de voorgaande vraaggesprekken, wil ik benadrukken dat het gesprek is gericht op weergave van je persoonlijke mening en ervaringen. Verkeerde antwoorden zijn dus ***niet*** mogelijk Er vindt wederom ***geen*** discussie plaats.

Het is belangrijk dat je in de gelegenheid wordt gesteld alles te zeggen wat je over de toets kwijt wilt. We kunnen starten met het gesprek als er geen vragen zijn.

A. Motivering

- Waarom is de keuze gevallen op de toetsen die wij vandaag bespreken?
- (Terughalen uitspraken vorige interview)

B. Validiteit: kennis, vaardigheden en inzicht

- Wat **meet** de toets?

Doorvragen

- Wat moeten leerlingen **weten** om de toets met succes te maken?
- Wat moeten leerlingen **kunnen** op het gebied van de Engelse taal?

- Hoe **zelfstandig** moet de leerling **kunnen leren** om de toets met succes te maken?
- In hoeverre meet je **inzicht** van de leerlingen?

C. Constructie en gebruik

- Hoe is de toets tot stand gekomen?
- Hoe kwam je aan de kennis en vaardigheden om deze toets te maken en te beoordelen?
- Hoe zijn de leerlingen op de toets voorbereid?
- Wat waren de afnamecondities?
- Hoe heb je gescoord en genormeerd: in detail Wanneer haal je de maximale score per vraag? Welke vragen zijn het eenvoudigst om te scoren, welke het moeilijkst? Welke vragen worden goed gemaakt, welke minder?
- Beschrijf teruggave, c.q nabespreking toets.
- Wat verwacht je dat leerlingen met de toets doen als je hem teruggeeft?

D. C.T.O

- In hoeverre bevordert voorbereiding, afname en nabespreking het luisteren, spreken, lezen en schrijven van de leerlingen in de Engelse taal?

E. Z.L.

- In hoeverre zet de voorbereiding, afname en teruggave, c.q. nabespreking van de toets aan tot zelfstandig en/of zelfverantwoordelijk leren?

APPENDIX VI: THE ORGINAL VERSIONS OF THE LANGUAGE TESTS (PRO MEMORI)

Little boy growing old before his time

AT just four years old, Elliott Lister ...1... to lose his memory and is wobbly on his feet.

In a few years' time he ...2... dead unless doctors can come up with a cure for a rare genetic condition.

Elliott ...3... the early stages of dementia - one of the symptoms of Niemann-Pick disease.

Sufferers are unable to deal with cholesterol in the body and the build-up in the liver, spleen and brain is eventually fatal. For a child to develop the disease, both parents must ...4... an abnormal gene - a chance estimated at three times less likely than winning the lottery jackpot. Unfortunately for Elliott, he is one of the unlucky few.

However, his parents Mark and Kerry ...5... to give up hope. They and other relatives ...6... £10,000 to send the boy to an American group working on drug therapies for the disease.

Elliott ...7... at eight months old after being ill from birth. Niemann-Pick usually ...8... school-age children, but can strike at any time. Life expectancy normally ...9... between five and 40 years.

Elliott's father Mark said: 'The doctors initially ...10... him only 18 months to live, but ...11... them so far.We ...12... not to give up hope.

'He is only at the development stage for an 18-month-old and will never catch up. He ...13... words he ...14... last year and he ...15... to stumble a lot. It's distressing because as yet there is no cure and we have to watch Elliott going backwards before he's even gone forwards.'

Mr Lister, 33, and his 30-year-old wife, from Oxted, -Surrey, ...16... two daughters - Briony, eight, and Sophie, six. Neither ...17... any signs of developing the condition so far.

Mr Lister, an airport maintenance worker, said: 'When we found out the implications of Elliott's disease, we ...18... , "Why us, what we ...19... to deserve this?"

'We ...20... a choice but to cope, there's nothing we can do to make it go away, so we have got to make the best of a bad thing. Kerry ...21... it better than I do, but it's not a matter of coping, all we can do is get on with it.

'We're hoping there ...22... be a cure, there's always that chance.

We try not to treat Elliott any differently to his sisters, but he does need special care, because he's a baby in a child's body. He's still in nappies. He gets frustrated and ...23...

Up till now Elliott ...24... to the National Institute of Health in Washington every year. His grandmother Barbara Lister, from Leeds, said: 'We are desperately raising funds for Elliott because sending him to America is our only hope - time ...25... out. He is a lovely little boy but he already ...26... to tire very easily now, which is devastating to watch.

'We ...27... to shut out the fact that he is not expected to live into his teens. We ...28... for a miracle, though common sense tells you that ...29... . All we can do is hope.

1. Start
2. Be
3. Suffer
4. Carry
5. Refuse
6. Raise
7. Be diagnosed
8. Affect
9. Vary
10. Give
11. Defy
12. Be determined
13. Forget
14. Learn
15. Start
16. Have
17. Show
18. Think
19. Do
20. Not have
21. Handle
22. Can
23. Not understand
24. Go
25. Run
26. Start
27. Try
29. Look
29. Not happen

By **HELEN GIBSON** BELFAST

IT WAS THE FOURTH DEADLINE IN NINE months, but this time, said British Prime Minister Tony Blair, it was absolute, with no plan B to fall back on. There would be "no peace at all" in Northern Ireland, he said, if this deadline was missed. With these dire warnings, Blair hoped to concentrate the minds of Northern Ireland's warring politicians, who convened last week in the drab Castle Buildings block in Belfast to try to save the deadlocked Good Friday Agreement, signed in the very same building just 14 months ago. "The entire civilized world will not understand if we can't put this together and make it work," Blair intoned.

In the end, something was put together, but it was far from clear that it will work. Five exhausting days of talks chaired by Blair and Irish Premier Bertie Ahern—with encouraging phone interventions from U.S. President Bill Clinton—failed to achieve a breakthrough. But there was no breakdown either. Paul Bew, professor of politics at Belfast's Queen's University, gave a 50-50 chance that a real deal was in the offing, one that might resolve the issue that has bedeviled the peace process since 1994: the surrender of weapons by illegal republican and loyalist paramilitary organizations.

Although the two main protagonists, David Trimble's Ulster Unionists and Gerry Adams' Sinn Fein, could not finally do a deal then and there, the two premiers drafted a document that outlined a way forward while addressing both parties' concerns and offering inducements to both sides. If accepted, the plan could break the deadlock. The premiers gave Adams and Trimble time to go back and sell the proposals to their constituents, away from the hothouse pressure of the talks. Analysts speculate that while Adams may have some problems with his supporters, Sinn Fein—and the I.R.A.—will ultimately agree. Trimble, on the other hand, who must constantly appease his own hardline supporters, may well ditch the plan.

The Blair-Aherne initiative calls for an executive with devolved powers that includes seats for republicans, nationalists and unionists to be nominated on July 15. It would take office three days later. Decommissioning would then start a few days after that and must be completed by next May, as originally specified in the Good Friday Agreement. The governments also offered a "failsafe" clause, backed by legislation, which promised to suspend the entire Good Friday Agreement—and all the institutions that go with it—if decommissioning or devolution fail to take place.

Decommissioning is still the main sticking point. The Ulster Unionists, Northern Ireland's largest party, had tried to get it started for years, but the issue has been so contentious that it has always been fudged. For the hard core of the I.R.A., handing over weapons is tantamount to surrender and has therefore been unacceptable. The present deadlock can be traced back to last October, when Trimble's party refused to allow Sinn Fein, the Irish Republican Army's political wing, to take up its allotted two seats in the power-sharing executive until the I.R.A. disarmed. No guns, no government, was Trimble's stance, holding that no democratic administration could be expected to include ministers who had a private army behind them. Sinn Fein, which has always maintained that it does not speak for the I.R.A—although Blair has insisted they are inextricably linked—pointed out that the Good Friday Agreement had not made decommissioning a precondition for government. No government, no guns, was the gist of Sinn Fein's rejoinder.

In fact, the Good Friday Agreement had indeed only spoken of parties "using their influence" to secure a handover, a fudge that caused Trimble dismay at the time. He only signed the accord after being reassured by a last minute letter from Blair saying that he believed "the process of decommissioning should begin straight away."

Politicians keep telling us we must build up trust with each other... If this fails, things will backslide. It makes one despair.

As last week's talks ground on, with roller-coaster mood swings of gloom and hope, the Wednesday midnight deadline was stretched until 4 a.m., with negotiations resuming the next day. An exhausted Ahern complained, "All these issues have been debated to death." But on Thursday, suddenly, Sinn Fein offered a concession that Blair described as a "seismic shift." A Sinn Fein paper said the leaders believed they "could succeed in persuading those with arms to decommission them in accordance with the agreement," as long as they got their seats on the executive. Although the Unionists said they found little new in the offer and were not impressed with the use of the aspirational word "could," the statement was a significant shift from Sinn Fein's previous point-blank refusal to make any link between the I.R.A.'s arms retension and their right to hold office.

At the same time, retired Canadian General John de Chastelain, the head of the commission on decommissioning, gave Sinn Fein a boost with his long-awaited assessment of the I.R.A.'s and other paramilitary bodies' intentions. He reported that although neither the I.R.A., nor any loyalist terrorist group except for one tiny splinter faction, had accepted the proposal to start decommissioning, the latest Sinn Fein statement "offers promise that decommissioning by all paramilitary groups may now begin."

Nationalists and republicans hailed the Sinn Fein statement and de Chastelain's report as a major breakthrough and a reason for the Unionists to sign up immediately. But Trimble's party questioned how Sinn Fein, which had always said it could not speak for the I.R.A., was now able to guarantee disarmament. The Unionists wanted guarantees from the I.R.A. itself. A furious Adams lashed out that the Unionists were simply not prepared "to have Catholics in government."

The peace process suddenly looked as fragile—and as urgent—as ever. While the politicians were talking, some 50 km away a huge military operation was launched outside the village of Drumcree near the town of Portadown, a sectarian tinderbox throughout its history. Soldiers unrolled miles of razor and barbed wire, dug ditches and ploughed up strips of land, turning the landscape into a battlefield scene. The British army was ensuring that a controversial protestant Orange Order parade did not ignore a ban prohibiting it from marching along a route that passed through a nationalist neighborhood. For years the march has triggered violent confrontations, culminating last year in the deaths of three young Catholic boys who were burnt alive in a petrol bomb attack on their home.

It is to put an end to this kind of horror that Northern Ireland's citizens voted overwhelmingly for the Good Friday Agreement in April last year. As Mrs. Phil McCullagh, 44, who carried a banner outside the talks reading "Build Trust Now," said: "The politicians keep telling us we must build up trust with each other. It's time they did it, too. If this fails, things will backslide. It makes one despair."

Such heartfelt concerns will no doubt be weighing heavily on Trimble as he considers his party's options over the next couple of weeks. ■

CLOCKWISE FROM TOP: FERMAN PARADES REUTERS; PETER MORRISON AP; AP; FERMAN PARADES REUTERS

H/V

Tekstverklaring - 4 ~~Havo~~ - April 2000

Alinea 1

1) In de tekst in sprake van een ernstige waarschuwing, gegeven door Premier Tony Blair. (5pt)
   a) Voor wie is de waarschuwing bedoeld?
   b) Waarvoor waarschuwt hij?

2) 'No plan B to fall back on' – Wat betekent dat? (10pt)

Alinea 2

3) Over welke belangrijke voorwaarde, omschreven als de meest cruciale basis voor het slagen van het vredesproces, wordt hier gesproken? (10pt)

4) Wat denkt Helen Gibson, de verslaggeefster, over de besprekingen in deze alinea? (10pt)

Alinea 3

5) David Trimble en Gerry Adams worden weggestuurd om te onderhandelen met hun achterban. Wie van deze twee zal naar Helen Gibson verwacht, de meeste tegenstand ondervinden? (10pt)

Alinea 5

6) 'Decommissioning' – Wat is dat? (10pt)

7) David Trimble zegt: 'No guns, no government' (10pt)
Gerry Adams zegt: 'No government, no guns'
Wat is het verschil?

Alinea 7

8) Wat is de 'seismic shift', zoals Blair dat zo mooi zegt? (10pt)

9) Waarom vallen Trimble's Unionisten over het woord 'could'? (5pt)

Alinea 10

10) 'The peace process suddenly looked .....' Waar slaat 'suddenly' op? (5pt)

11) Wat was het doel van de 'military operation' in Drumcree? (5pt)

Alinea 12

12) Volgens de conclusie van het artikel ligt de toekomst van Noord Ierland in de handen van Trimble. Verklaar waarom. (10pt)

A DAY'S WAIT

by E.Hemingway

He came into the room to shut the windows while we were still in bed and I saw he looked ill. He was shivering, his face was white, and he walked slowly as though it ached to move.

'What's the matter, Schatz?'

'I've got a headache.'

'You better go back to bed.'

'No, I'm all right.'

'You go to bed. I'll see you when I'm dressed.'

But when I came downstairs he was dressed, sitting by the fire, looking a very sick and miserable boy of nine years. When I put my hand on his forehead I knew he had a fever.

'You go up to bed,' I said, 'you're sick.'

'I'm all right,' he said.

When the doctor came he took the boy's temperature.

'What is it?' I asked him.

'One hundred and two.'

Downstairs, the doctor left three different medicines in different coloured capsules with instructions for giving them. One was to bring down the fever, another a purgative, the third to overcome an acid condition. The germs of influenza can only exist in an acid condition, he explained. He seemed to know all about influenza and said there was nothing to worry about if the fever did not go above one hundred and four degrees. This was a light epidemic of flue and there was no danger if you avoided pneumonia.

Back in the room I wrote the boy's temperature down and made a note of the time to give the various capsules.

'Do you want me to read to you?'

'All right. If you want to,' said the boy. His face was very white and there were dark areas under his eyes. He lay still in the bed and seemed very detached from what was going on.

I read aloud from Howard Pyle's Book of Pirates; but I could see he was not following what I was reading.

'How do you feel, Schatz?' I asked him.

'Just the same, so far,' he said.

I sat at the foot of the bed and read to myself while I waited for it to be time to give another capsule. It would have been natural for him to go to sleep, but when I looked up he was looking at the foot of the bed, looking very strangely.

'Why don't you try to go to sleep? I'll wake you up for the medicine.'

'I'd rather stay awake.'

After a while he said to me, 'You don't have to stay in here with me, Papa, if it bothers you.'

'It doesn't bother me.'

'No, I mean you don't have to stay if it's going to bother you.'

I thought perhaps he was a little lightheaded and after giving him the prescribed capsules at eleven o'clock I went out for a while.

It was a bright, cold day, the ground covered with a sleet that had frozen so that it seemed as if all the bare trees, the bushes, the cut brush and all the grass and the bare ground had been varnished with ice. I took the young Irish setter for a little walk up the road and along a frozen creek, but it was difficult to stand or walk on the glassy surface and the red dog slipped and slithered and I fell twice, hard, once dropping my gun and having it slide away over the ice.

We flushed a covey of quail under a high bank with overhanging brush and I killed two as they went out of sight over the top of the bank. Some of the covey lit in trees, but most of them scattered into brush piles and it was necessary to jump on the ice-coated mounds of brush several times before they would flush. Coming out while you were poised

unsteadily on the icy, springy brush they made difficult shooting and I killed two, missed five, and started back pleased to have found a covey close to the house and happy there were so many left to find on another day.

At the house they said the boy had refused to let anyone come into the room.

'You can't come in,' he said, 'You mustn't get what I have.'

I went up to him and found him in exactly the position I had left him, white-faced, but with the tops of his cheeks flushed by the fever, staring still, as he had stared, at the foot of the bed.

I took his temperature.

'What is it?'

'Something like a hundred,' I said. It was one hundred and two and four-tenths.

'It was a hundred and two,' he said.

'Who said so?'

'The doctor.'

'Your temperature is all right,' I said. 'It's nothing to worry about.'

'I don't worry,' he said, 'but I can't keep from thinking.'

'Don't think,' I said. 'Just take it easy.'

'I'm taking it easy,' he said, and looked straight ahead. He was evidently holding tight on to himself about something.

'Take this with water.'

'Do you think it will do any good?'

'Of course it will.'

I sat down and opened the Pirate book and commenced to read, but I could see he was not following, so I stopped.

'About what time do you think I'm going to die?' he asked.

'What?'

'About how long will it be before I die?'

'You aren't going to die. What's the matter with you?'

'Oh, yes, I am. I heard him say a hundred and two.'

'People don't die with a fever of one hundred and two. That's a silly way to talk.'

'I know they do. At school in France the boys told me you can't live with forty-four degrees. I've got a hundred and two.'

He had been waiting to die all day, ever since nine o'clock in the morning.

'You poor Schatz,' I said. 'Poor old Schatz. It's like miles and kilometres. You aren't going to die. That's a different thermometer. On that thermometer thirty-seven is normal. On this kind it's ninety-eight.'

'Absolutely,' I said. 'It's like miles and kilometres. You know, like how many kilometres we make when we do seventy miles in the car?'

'Oh,' he said.

But his gaze at the foot of the bed relaxed slowly. The hold over himself relaxed too, finally, and the next day it was very slack and he cried very easily at little things that were of no importance.

- - -

A DAY'S WAIT by Ernest Hemingway

varnished with ice (ll. 49 - 50) : met een vernislaagje van ijs
a covey of quail (l. 54) : a flock of birds (kwartels)
brush (l. 56) : kreupelhout
to flush (l. 58) : to fly up suddenly
poised (l. 58) : balanced

ANSWER IN CORRECT ENGLISH:

1) What is the story about? Give a brief outline.

2) What is the theme of the story?

3) Describe the boy's character; what do you know about the boy?

4) Father - son relationship; what can you say about it?

5) What is the crucial line of the story? Why?

6) "You don't have to stay in here with me, Papa, if it bothers you." (ll. 41-2)
What is "it"?

7) Explain the last line of the story; ".. he cried very easily at little things that were of no importance." (l. 106)

8) Construction of the story. The story can be divided into three parts.
a) When does part two begin? When does part three begin? Mention the numbers of the lines.
b) What is the function of part two?

9) Explain the title.

10) Do you think this is a good story? Why (not)? Personal valuation.

UNICOM FINALS 4 VWO //////// UNIT EIGHT

A

I. Give the main forms of:

1) vasthouden 2) schudden 3) leggen 4) vliegen 5) onderwijzen 6) vallen

II. Translate into correct English:

1) Ik neem aan dat de cursusgelden inclusief BTW zijn.
2) Hierbij stuur ik een foto van mezelf en een geadresseerde envelop met een internationale antwoordcoupon.
3) Ik hoop dat mijn brief u bereikt voor de sluitingsdatum voor sollicitaties ondanks de huidige poststaking.
4) Gefeliciteerd met je fantastische prestatie, ook namens de rest van de familie.
5) Ze heeft een donkere gelaatskleur en prachtige tanden.

III. Translate into correct Dutch:

1) You deserve a pat on the back.
2) RSVP
3) She is always very smartly dressed.
4) Let's get to the heart of the matter.
5) I think I am well suited for this job.

IV. Choose the correct form:

1) It smells awful/awfully here.
2) The performance was clever/cleverly done.
3) The losses of the film appeared considerable/considerably.
4) The pupil behaved extreme/extremely bad/badly.
5) Did they steal your wallet? How low/lowly can one get!

V. Translate into English:

1) De betrokkenen konden het nauwelijks geloven.
2) Zijn stem klonk erg ongeduldig en hij was duidelijk zenuwachtig.
3) Die leraar vertelt zijn verhalen altijd heel levendig.
4) Wij kunnen onmogelijk betalen; we hebben bijna geen geld meer.
5) De aanwezigen waren uitermate teleurgesteld.

B

LETTER WRITING

Translate this letter into correct English:

Hallo Daphne,

Ik heb op de BBC een programma gezien over hondengevechten. Het was afschuwelijk! Ik wist niet dat dit mogelijk was in een beschaafd land als Groot-Brittannië. Maar misschien gebeurt het ook in Nederland. De honden die gebruikt werden waren voornamelijk pitbull terriërs. Ze worden gefokt omdat ze agressief en sterk zijn. Het is ongelooflijk maar ik hoorde dat deze honden getraind worden om elkaar aan stukken te scheuren. Soms duren deze gevechten langer dan een uur. De eigenaren en toeschouwers schijnen van deze gevechten te genieten. Wat een verachtelijke mensen! Na een gevecht durven de eigenaren niet naar een dierenarts te gaan. Ze zijn bang dat de dierenarts hen zal aangeven bij de politie. Daarom krijgen de dieren vaak een slechte en onprofessionele behandeling. Soms worden ze zelfs afgemaakt.
Vreselijk, hè? Ben jij lid van de RSPCA? Ik weet zeker dat ik lid word van de Nederlandse dierenbescherming. Deze misdaden moeten gestopt worden.

Hartelijke groeten,

Online news audience becomes more mainstream
Major newspaper sites lose audience share
By David Noack

1 Weather information sites, along with entertainment and local news, are among the most popular online destinations as the Web continues to become more mainstream, says a Pew Center study released Thursday, Jan. 14, in **1** ___ . Meanwhile, major newspapers' Web sites have lost some online audience share to broadcast TV sites. The findings by the Pew Research Center for the People and the Press show that in 1996, when just 23% of Americans went online for news, technology stories topped the preference list. Today, with 41 percent of all American adults online, **2** ___ is the most popular attraction.

2 This comes as personal computers and online services continue to make inroads into the home, while going online is becoming easier and **3** ___ . For online news organizations, there is good news in that the numbers of all people who get news online at least weekly continues to grow, starting from 4% in 1995 to between 15% to 26% now, the study shows. Among strictly Internet users, the numbers getting news online weekly ranges from 37% to 64%. Poll officials said the fluctuations in the figures of people getting online news weekly is related to what may be happening in the news. Major or breaking news stories attract **4** ___ users. 'The number-one draw is now the weather,' says Andrew Kohut, the poll director. 'Sixty-four percent now say they go on the Internet for the weather, which is much more typical of American **5** ___. We no longer have the intense group of technological, professionally oriented people dominating the Internet.' He also cited local community news going from 27% percent in a 1996 study to 42% in 1998, and entertainment news going from 50% to 58% in the same time period. Technology news remains popular, falling to 59% from 64% of online users. 'What does this tell you?' asks Kohut. 'Weather, entertainment, local news all sound like the 6 o'clock news. As the Internet population becomes more **6** ___ , it will begin to look in its preferences, its habits and its behavior more like general news audiences.' National newspaper Web sites are being viewed by a lower percentage of Web users, according to the study. In 1995, 23% of all Net users visited **7** ___ sites. Only 16% did last year. Meanwhile, broadcast TV sites increased from 20% of online users in 1995 to 22% in 1998. MSNBC alone now enjoys readership from 11% of all online users. The survey results were gleaned from a nationwide poll of 3,184 adults conducted throughout the fall of 1998. There are roughly 74 million Internet users in the United States. Kohut says the survey shows that new Internet users, people who have gone online over the last year, are displaying different information, news and usage patterns than people who have been connected for more than a year. Middle-aged and middle-income people, both men and women, are using the Internet in increased numbers, as are people with less education than a college degree. Among Internet newcomers, 39% never attended college, and 23% have household incomes below $30,000 a year. 'When we asked the respondents, mostly what do you use the Internet for -- work or pleasure, among the new users, 52% say pleasure and 24% say for work ... So we have quite a different **8** ___,' Kohut observes. 'We see the Internet changing and this population, for all intents and purposes is, to use our phrase, "normalizing" the Web,' he says. As the number of people who go online for news continues to grow, online information remains a **9** ___, rather than a replacement, for traditional print and broadcast news. That said, for a small group of users, there are indications that the Internet is beginning to erode their traditional news sources. One-fifth of Internet users (21%) regularly read news stories online instead of reading them in a newspaper or watching them on television and 16% say they get more news from **10** ___ than from print or broadcast. People are using the Internet to garner news because of its accessibility, convenience and the ability to conduct searches. Forty-six percent say access to information that is not easily available elsewhere is a key reason for using the Web. Internet users read newspapers and listen to radio news at higher rates than do non-users. Sixty-four percent watched TV news the day before the survey interview, 51% read a newspaper and 48% listened to radio news. These usage rates are **11** ___ than for the non-Internet using population. Customized news doesn't appear to be as popular among online newcomers. While 18% of older Internet users get customized news updates and 23% get stories e-mailed to them, only 12% of new users have stories e-mailed to them and 16% get customized updates. In November, Jupiter Communications, a New York-based online and technology research firm, released a study showing that American consumers of Web news trust online news as much as they do news from newspapers, radio, television and magazines. The Pew study also finds that 49% of Web users say that Internet news is 'more accurate than news found in traditional print and broadcasts outlets.' But only 28% of non-online news users agree with that assessment. Also, 55% of Americans do not think that news organizations' Web sites are any more or less accurate than information from **12** ___ sources. In addition to online news, the study also addresses a number of other Internet-related issues, such as basic patterns of Internet use, how Americans used the Internet in the 1998 elections, the attitudes, beliefs and behaviors of Internet users and American's attitudes toward the Internet and technology. For more information about the study, visit http://www.people-press.org.

Choose the correct alternatives to fill the gaps

1
A London
B Washington

2
A the weather
B science

3
A cheaper
B more expensive

4
A less
B more

5
A consumers
B newspapers

6
A educated
B mainstream

7
A local paper
B national paper

8
A answer
B population

9
A necessity
B supplement

10
A newspapers
B online sources

11
A higher
B lower

12
A Internet
B traditional

HANDELINGSDEEL ENGELS KLAS 4 /////// WERKSTUK LITERATUUR

In groepjes van 4 (eventueel 3) behandel je de roman die je gekozen hebt. Je kunt dit doen op de manier die je als groepje het beste vindt, maar aan te raden is om het leeswerk **thuis** te doen en de beschikbare lessen te gebruiken om de gelezen hoofdstukken te bespreken en daadwerkelijk bezig te zijn met de onderstaande delen van de opdracht.

OPDRACHT

Maak een werkstuk over het boek dat je gelezen hebt. In dit werkstuk geef je in (kleine) hoofdstukken informatie over de volgende aspecten:

- summary — geef een samenvatting van het verhaal in maximaal 500 woorden
- title — leg uit waarom de titel belangrijk is; vermeld op welke manier de titel met het verhaal te maken heeft
- setting — beschrijf waar het verhaal zich afspeelt en maak duidelijk in hoeverre de setting een rol speelt in het verhaal
- theme — wat is het thema van het verhaal? Leg uit waarom dat zo is
- main character — beschrijf de hoofdfiguur en geef aan of hij/zij verandert in de loop van het verhaal; zo ja, maak dan die ontwikkeling duidelijk
- atmosphere — geef informatie over de sfeer van het verhaal
- quotation — kies één bepaalde scene/alinea/zin en leg uit waarom je dit citaat belangrijk vindt
- question — bedenk een bepaalde vraag over het boek en geef een uitvoerig antwoord op die vraag
- personal valuation — ieder groepslid geeft zijn/haar persoonlijke mening over het boek en geeft commentaar op de manier van (samen)werken

AANPAK

Elk groepje kiest een aanvoerder die namens zijn of haar groepje verantwoordelijk is. De deadline (de uiterste datum waarop het project ingeleverd moet zijn) hoor je van je docent. Zorg dat elk groepslid een copie heeft van het in te leveren project. De groepsleden krijgen in principe allemaal dezelfde beoordeling. Goed overleg is dus noodzakelijk! Spreek met elkaar af hoe je de zaak gaat aanpakken. Maak een goede planning! Beslis wat je per les gaat lezen/behandelen. Wat doe je thuis, wat doe je in de klas? De aanvoerder ziet erop toe dat ieder groepslid zijn/haar afspraken nakomt. In een schrift houdt de aanvoerder bij:

1) wat er elke les gedaan is
2) wat er afgesproken is voor de volgende les(sen).

Ook kunnen er vragen aan de docent gesteld worden die de volgende les met hem/haar in je groepje besproken kunnen worden. Aan het eind van elke les levert de aanvoerder het schrift in bij de docent, zodat die van jullie voortgang en vragen op de hoogte blijft.
Het werkstuk dient in het Engels geschreven te zijn.

BEOORDELING WERKSTUK KLAS 4

Namen groepsleden: 1) (aanvoerder)
2)
3)
4)

Titel van het boek:

VERZORGING (10) :
(netheid, indeling, illustraties, enz.)

PRESENTATIE (10) :
(taalgebruik, stijl)

SAMENWERKING (10) :
(per leerling)

SUMMARY (10) :

TITLE (5) :

SETTING (10) :

THEME (5) :

MAIN CHARACTER (10) :

ATMOSPHERE (10) :

QUOTATION (10) :

QUESTION (10) :

(tussen haakjes het maximaal aantal te verdienen punten)

OPMERKING: **TOTAALSCORE**:

REFERENCES

Alabau, I., Bonnet, G, de Bot, K., Bramsby, J., Dauphin, L., Erickson, G., et al. (2004). *The assessment of pupils' skills in English in eight European countries 2002*. Paris: European Network of policy makers for the evaluation of education systems.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Aljaafreh, A., & Lantolf, J. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal, 78*, 465-483.

American Psychological Association (APA). (1954). *Standards for educational and psychological tests and manuals*. Washington, DC: APA.

Anderson, J. R., & Bower, G. H. (1980). *Human associative memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, R. C., & Ausubel, D. P. (1965). *Readings in the psychology of cognition*. New York: Holt, Rinehart and Winston.

Antón, M. (2003). *Dynamic assessment of advanced foreign language learners*. Paper presented at the annual conference of the American Association of Applied Linguistics (AAAL).

Argyris, C., & Schön, D. A. (1974). *Theory in practice: Increasing professional effectiveness.* San Francisco: Jossey-Bass Publishers.

Arnold, J. (Ed.). (1999). *Affect in language learning*. Cambridge: Cambridge University Press.

Aronson, E., & Lindzey, G. (1985). *Handbook of social psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates

Association of Language Testers in Europe (ALTE). (1998). *Multilingual glossary of language testing terms* (Vol. 6). Cambridge: Cambridge University Press.

Ausubel, D. A. (1963). Cognitive structure and the facilitation of meaningful verbal learning. *Journal of Teacher Education, 14*, 217-221.

Ausubel, D. A. (1968). *Educational psychology: A cognitive view. New York: Holt, Rinehart and Winston*. New York: Holt, Rinehart and Winston.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bailey, K. M. (1986). *Class lecture, spring 1986.*: Monterey Institute of International Studies.

Baker, D. (1989). *Language testing. A critical survey and practical guide*. London: Edward Arnold.

Bandura, A. (1971). *Social learning theory*. New York: General Learning Press.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall.

Bardovi-Harlig, K., & Hartford, B. (1993). Learning the rules of academic talk: A longitudinal study of pragmatic development. *Studies in Second Language Acquisition, 15*, 279-304.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bausch, K., Christ, H., & Königs, F. G. (Eds.). (2002). *Der Gemeinsame europäische Referenzrahmen für Sprachen in der Diskussion*. Tübingen: Günter Narr Verlag.

Beck, C. (1993). Postmodernism, pedagogy, and philosophy of education [Electronic Version]. Retrieved 28 May 2006 from http://www.ed.uiuc.edu/eps/PES-Yearbook/93_docs/BECK.HTM.

Benson, P. (1997). The philosophy and politics of learner autonomy. In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 18-34). New York: Longman.

Benson, P., & Voller, P. (1997). *Autonomy and independence in language learning*. New York: Longman.

Berns, M. S. (1990). *Contexts of competence: Social and cultural considerations in communicative language teaching*. New York: Plenum Press.

Billmyer, K. (1990). "I really like your lifestyle": ESL learners learning how to compliment. *Penn Working Papers in Educational Linguistics, 6*(2), 31-48.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart.

Bloomfield, L. (1942). *Outline guide for the practical study of foreign languages*. Baltimore, Md: Linguistic Society of America.

Boekaerts, M., & Simons, P. J. R. (1995). *Leren en instructie: Psychologie van de leerling en het leerproces*. Assen: Van Gorcum.

Breen, M. P. (1985). Authenticity in the language classroom. *Applied Linguistics, 6*, 60-70.

Brooks, J. G., & Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms*. Alexandria: Association for Supervision and Curriculum Development.

Brooks, N. (1967).

Brown, H. D. (1972). Cognitive pruning and second language acquisition. *Modern Language Journal, 56*, 218-222.

Brown, H. D. (2000). *Principles of language learning and teaching* (4th ed.). New York: Addison Wesley Longman.

Brown, H. D. (2001). *Teaching by principles* (2nd ed.). New York Addison Wesley Longman.

Bruner, J. (1996). *Toward a theory of instruction* Cambridge, MA: Harvard University Press.

Budoff, M., & Friedman, M. (1964). "Learning Potential" as an assessment approach to the adolescent mentally retarded. *Journal of Consulting Psychology, 28*, 434-439.

Buss, S. (2002). Personal autonomy. *The Stanford Encyclopedia of Philosophy* Retrieved 14th March, 2006, from http://plato.stanford.edu/archives/win2002/entries/personal-autonomy

Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grade on intrinsic instruction and performance. *Journal of Educational Psychology, 78*, 210-216.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters Ltd.

Byram, M., & Zarate, G. (1997). *The sociocultural and intercultural dimension of language learning and teaching*. Strasbourg: Council of Europe.

Canady, R. L., & Hotchkiss, P. R. (1989). It's a good score! Just a bad grade. *Phi Delta Kappan, 71*(1), 68-71.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt (Eds.), *Language and communication*. London: Longman.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*(1), 1-47.

Carroll, J. B. (1961). *Fundamental considerations in testing for English language proficiency of foreign students* Washington, DC: Center for Applied Linguistics.

Carter, K., & Doyle, W. (1995). Preconceptions in learning to teach. *The Educational Forum, 59*(2), 186-195.

Carter, K., & Doyle, W. (1995). Teacher-researcher relationship in the study of teaching and teacher education. *Peabody Journal of Education, 70*(2), 162-174.

Chapelle, C., & Abraham, R. G. (1990). Cloze method: What difference does it make? *Language Testing, 7*, 121-146.
Chapelle, C., & Green, P. (1992). Field independence/dependence in second language acquisition research. *Language Learning, 42*, 47-83.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Cobb, P., Confrey, J., deSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Journal of the Learning Sciences, 2*(2), 141-178.
Cohen, A. (1998). *Strategies in learning and using a second language*. New York: Addison Wesley Longman.
Costa, A., & Liebmann, R. (1995). Process is as important as content. *Educational Leadership, 52*(6), 23-24.
Cotterall, S. (1995). Readiness for autonomy: Investigating learner beliefs. *System, 23*(2), 195-205.
Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
Council of Europe, Council for Cultural Co-operation, & Education Committee. (1998). *Modern Languages: Learning, Teaching Assessment. A Common European Framework of Reference for Languages*. from http://www.coe.int/T/DG4/Linguistic/CADRE_EN.asp.
Covington, M. V., & Omelich, C. L. (1984). Task-oriented versus competitive learning structures: motivational and performance consequences. *Journal of Educational Psychology, 76*, 1199-1213.
Crabbe, D., Hoffmann, A., & Cotterall, S. (2001). Examining the discourse of learner advisory sessions. *AILA Review, 15*(Dam, L. (ed.) Learner autonomy: new insights.), 2-15.
Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.
Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*, 367-383.
Crookes, G., & Schmidt, R. W. (1991). Motivation: Reopening the research agenda. *Language Learning, 41*(4), 469-512.
Curry, M. J. (2006). Action research for preparing reflective language teachers. *TESOL HEIS (Higher education interest section), 25*(1), also accessed June 2008: www.sfu.ca/heis/archive/2025-2001Curry_ActionResearch_eddone.doc.
d'Anglejan, A., Harley, B., & Shapson, S. (1990). Student evaluation in a multidimensional core French curriculum. *The Canadian Modern Language Review, 47*, 107-124.
Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. London: Routledge Falmer.
Darian, S. (1969). Backgrounds of modern language teaching: Sweet, Jespersen, and Palmer. *The Modern Language Journal, 53*, 545-550.
Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 64*(1), 5-30.
Davidson, F., & North, B. (2006). Flexibility is proof of a good 'Framework". How to bend the rules of test writing with Europe's guide to language ability [Electronic Version]. Retrieved July 2007 from http://education.guardian.co.uk/tefl/teaching/story/0,,1950500,00.html.
Davies, A. (1997). Demands of being professional in language testing. *Language Testing, 14*(3), 328-339.
DeCharms, R. (1984). Motivation enhancement in educational settings. *Research on Motivation in Education, 1*, 275-310.
Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum Press.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Berlin: Springer.
Dickinson, L. (1995). Autonomy and motivation. A literature review. *System, 23*(2), 165-174.
Donato, R. (1994). Collective scaffolding in second language learning. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second Language research*. Norwood, N.J.: Ablex.
Dörnyei, Z. (2002). *Motivational determinants of the quality and quantity of student performance in communicative language tasks*. Paper presented at the annual conference of the American Association of Applied Linguistics (AAAL).
Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.
Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.
Edelenbos, P., van der Schoot, F., & Verstralen, H. (2000). *Balans van het Engels aan het einde van de basisschool 2: Uitkomsten van de tweede peiling in 1996*. Arnhem: CITO
Edelhoff, C. (1983). Kommunikative Grundlagen des Englishunterrichts. In H. Christ & M. J. Legutke (Eds.), *Fremde Texte Verstehen: Festschrift für Lothar Bredella* (pp. 40-50). Tübingen: Gunter Narr Verlag.
Ehrman, M. (1996). *Understanding second language learning difficulties*. Thousand Oaks, Ca.: Sage Publications.
Ehrman, M., & Oxford, R. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal, 74*, 311-327.
Ehrman, M., & Oxford, R. (1995). Cognition plus: Correlates of language proficiency. *Modern Language Journal, 79*, 67-89.
Ellis, N., & Laporte, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. In A. Groot de & J. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*. Mahwah NJ: Lawrence Erlbaum Associates.
Ellis, R. (1986). *Understanding second Language acquisition*. Oxford: Oxford University Press.
Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.
Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition, 24*, 223-236.
Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
Elsen, A., & St.John, O. (2003). Profiling for progress: Learner autonomy in writing skills development. In K. van Esch & O. St.John (Eds.), *A framework for freedom*. Frankfurt am Main: Peter Lang Verlag.
Elsen, A., & St.John, O. (2007). Learner autonomy and intercultural competence. In M. Jiménez-Raya & L. Sercu (Eds.), *Challenges in teacher development: Learner autonomy and intercultural competence*. Frankfurt an Main: Peter Lang Verlag.
Elsen, A., van Esch, K., Bohn, P., & Friedrich, M. (2003). Learning to teach how to read in a foreign language. In K. Van Esch & O. St.John (Eds.), *A framework for freedom*. Frankfurt am Main: Peter Lang GmbH.
Engelen, A. (2002). *Coaching binnenstebuiten: een onderzoek naar coaching van docenten door docenten.* Thesis Radboud University, Nijmegen.
Eurybase. (2003). The database on education systems in Europe (Publication. Retrieved 24th June 2003, from Council of Europe (Eurydice): http://www.eurydice.org/portal/page/portal/Eurydice/EuryPresentation
Farhady, H. (1980). *Justification, development, and validation of functional language testing.* University of California, Los Angeles.

Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment*. Baltimore, Md: University Park Press.
Feuerstein, R., Rand, Y., & Rynders, J. E. (1988). *Don't accept me as I am. Helping retarded performers excel*. New York: Plenum.
Feuerstein, R. L., Falik, L., Rand, Y., & Feuerstein, R. S. (2003). *Dynamic assessment of cognitive modifiability*. Jerusalem: ICELP Press.
Fischer, J. M., & Ravizza, M. (1993). *Perspectives on moral responsibility*. Ithaca, NY: Cornell University Press.
Flamini, E., & Jiménez-Raya, M. (2007). Action research: Professional development through enquiry. In M. Jiménez-Raya & L. Sercu (Eds.), *Challenges in teacher development: Learner autonomy and intercultural competence*. Frankfurt am Main: Peter Lang GmbH.
Flavell, J. H. (1976). Metacognitive aspects of problem solving. In B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Pittsburgh: University of Pittsburgh.
Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911.
Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics, 19*, 1-23.
Frawley, W., & Lantolf, J. P. (1985). Second language discourse: A Vygotskyan perpective. *Applied Linguistics, 6*, 19-44.
Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27.
Freire, P. (1972). *Pedagogy of the oppressed*. Harmondsworth: Penguin.
Fries, C. (1945). *Teaching and learning English as a foreign language*. Ann Arbor: University of Michigan Press.
Fries, C. (1952). *The structure of English*. New York: Harcourt, Brace, & World.
Fulcher, G. (2000). The ‘communicative’ legacy in language testing. *System, 28*(4), 483-497.
Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly: An International Journal, 1*(4), 253-266.
Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, Mass: Newbury House.
Gardner, R. C., & MacIntyre, P. D. (1991). An instrumental motivation in language study: Who says it isn't effective? *Studies in Second Language Acquisition, 13*, 57-72.
Gass, S., & Varonis, E. (1985). Task variation and native/non-native negotiation of meaning. In S. Gass & C. Madden (Eds.), *Input in second language acquisition*. Edinburgh: Edinburgh University Press.
Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah NJ: Lawrence Erlbaum Associates.
Giroux, H. A. (1981). *Ideology, culture and the process of schooling*. Philadelphia: Temple University Press.
Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Transaction.
Good, T. L., & Brophy, J. E. (1991). *Looking in classrooms*. New York: Harper Collins.
Grabois, H. (2004). What does SCT research reveal about second language learning? In K. Van Esch & O. St.John (Eds.), *New insights into foreign language teaching and learning* (Vol. 9, pp. 35-50). Frankfurt am Main: Peter Lang GmbH.
Groot de, A. M. B., & Kroll, J. F. (1997). *Tutorials in bilingualism: Psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates
Guberina, P. (1972). Sur la notion de 'structuro-global'. *Revue de Phonetique Appliquée, 21*, 9-13.

Guthke, J., Heinrich, A., & Caruso, M. (1986). The diagnostic program of "syntactical rule and vocabulary acquisition": A contribution to the psychodiagnosis of foreign language learning ability. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities*. Amsterdam: Elsevier.

Habermas, J. (1970). *Zur Logik der Sozialwissenschaften*. Frankfurt am Main: Suhrkamp Verlag.

Habermas, J. (1971). *Theorie und Praxis* (2nd ed.). Frankfurt am Main: Suhrkamp Verlag.

Halliday, M. (1970). Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics*. Harmondsworth: Penguin.

Halliday, M. (1973). *Explorations in the functions of language*. London: Edward Arnold Press.

Halliday, M. (1978). *Language as a social semiotic: The social interpretation of language and meaning*. Baltimore: University Park Press.

Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing, 14*(3), 295-303.

Handal, G., & Lauvas, P. (1987). *Promoting reflective teaching: Supervision in action*. Milton Keynes: SHRE and Open University.

Hanson, F. E. (1993). *Testing testing: Social consequences of the examined life*. Berkeley, CA: University of California Press.

Harley, B., & Swain, M. (1984). The interlanguage of immersion students and its implications for second language teaching. In A. Davies, C. Criper & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 291-311). Edinburgh: Edinburgh University Press.

Hartford, B. (1997). Sociolinguistics in language teacher education programs. In K. Bardovi-Harlig & B. Hartford (Eds.), *Beyond methods: Components of second language teacher education* New York: McGraw Hill.

Hatch, E. (1978). Discourse analysis and second language acquisition. In E. Hatch (Ed.), *Second language acquisition: A Book of readings*. Rowley, MA: Newbury House.

Hatch, E. (1992). *Discourse and language education*. Cambridge: Cambridge University Press.

Haywood, H. C., Brown, A. L., & Wingenfeld, S. (1990). Dynamic approaches to psycho-educational assessment. *School Psychology Review, 19*, 411-422.

Hermans-Nymark, L. (2006). *English in the classroom: Why not? Classroom discourse patterns and teachers' beliefs.* Thesis Radboud University, Nijmegen.

Hilgard, E. (1963). Motivation in learning theory. In S. Koch (Ed.), *Psychology: A study of science, volume 5*. New York: McGraw-Hill.

Hoffman, E. (1989). *Lost in translation. A life in a new language*. New York: Dutton.

Hoffman, S. Q. (1997). Field dependence/indepence in second language acquisition and implications for educators and instructional designers. *Foreign Language Annals, 30*, 221-234.

Holec, H. (1981). *Autonomy in foreign language learning*. Oxford: Pergamon Press.

Howatt, A. (1984). *A history of English language teaching*. Oxford: Oxford University Press.

Howatt, A. P. R., & Widdowson, H. G. (2004). *A history of English language teaching*. New York: Oxford University Press.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hull, C. L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts

Hunt, J. M. (1971). Toward a history of intrinsic motivation. In H. I. Day, D. E. Berlyne & D. E. Hunt (Eds.), *Intrinsic motivation: A new direction in education*. Ontario: Ontario Institute for Studies in Education.

Hymes, D. (1967). Why linguistics needs the sociologist. *Social Research, 34*, 632-647.

Hymes, D. (1971). Competence and performance in linguistic theory. In R. Huxley & E. Ingram (Eds.), *Language acquisition: Models and methods*. London: Academic Press.

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.

Illich, I. (1979). *Deschooling society*. Harmondsworth: Penguin.

Inspectie van het Onderwijs. (1999a). *Engels in de basisvorming: Evaluatie van de eerste vijf jaar*. Den Haag: SDU.

Inspectie van het Onderwijs. (1999b). *Werk aan de basis: Evaluatie van de basisvorming na vijf jaar*. Den Haag: SDU.

International Language Testing Association (ILTA). (2000). *Code of Ethics for Foreign/Second Language Testing*. Hong Kong: ILTA.

Jakobovits, L. (1968). The physiology and psychology of second language learning. In E. Birkmaier (Ed.), *Britannica review of foreign language education* (Vol. 1). Chicago: Encyclopaedia Brittanica, Inc.

Janssen, T., ten Dam, G., & van Hout-Wolters, B. (2002). *Vaardigheden voor zelfstandig leren: Een praktijkgericht overzicht van onderzoek*. Assen: Koninklijke Van Gorcum.

Jarvis, G. (1972). Teacher education goals: they're tearing up the street where I was born. *Foreign Language Annals, 6*, 198-205.

Jespersen, O. (1904). *How to teach a foreign language* (13th, 1967 ed.). London: Allen and Unwin.

Kamp, M. (2000). *Centrale concepten in het curriculum: Het voorbeeld homeostase in het curriculum biologie van de bovenbouw vwo.* Thesis Radboud University, Nijmegen.

Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.

Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research, 72*(2), 177.

Kanpol, B. (1994). *Critical pedagogy: An Introduction*. Westport CT: Bergin & Garvey.

Kanpol, B. (1994). *Issues and trends in critical pedagogy*. Creskill NJ: Hampton Press.

Kearsley, G. (2002). Genetic epistemology (J.Piaget). *Explorations in learning and instruction: The theory into practice database* Version 2.6 May 2006. Retrieved 14th June, 2008, from http://tip.psychology.org/piaget/html

Keefe, J. W. (1979). *Student learning styles: Diagnosing and prescribing programs*. Reston, VA: National Association of Secondary School Principals

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigelruth (Ed.), *Instructional design theories and models: An overview of their current status* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kelly, L. (1969). *25 centuries of language teaching*. Rowley, Mass.: Newbury house Publishers.

Kenny, B. (1993). For more autonomy. *System, 21*(4), 431-442.

Kinginger, C. (2004). Communicative foreign language teaching through telecollaboration. In K. Van Esch & O. St.John (Eds.), *New insights into foreign language learning and teaching* (Vol. 9, pp. 101-113). Frankfurt am Main: Peter Lang GmbH.

Kowal, M., & Swain, M. (1997). From semantic to syntactic processing. In K. Johnson & M. Swain (Eds.), *Immersion education: International perspectives*. Cambridge: Cambridge University Press.

Kozulin, A. (1998). *Psychological tools: A sociocultural approach to education*. Cambridge, Massachusetts: Harvard University Press.

Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension of at-risk students. *School Psychology International, 23*, 112-127.

Kramsch, C. (1993). *Context and culture in language teaching*. New York: Oxford University Press.
Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New York: Pergamon Press.
Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. New York: Longman.
Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly, 35*(4), 537-560.
Kunnan, A. J. (1999). Language testing: Fundamentals. In B. Spolsky (Ed.), *Concise encyclopedia of educational Linguistics* (pp. 707-711). Oxford: Elsevier Science Ltd.
Lado, R. (1949). *Measurement in English as a foreign language with special reference to Spanish-speaking adults.* University of Michigan., Ann Arbor.
Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.
Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman Group ltd.
Lantolf, J. P. (2004). Sociocultural theory and second and foreign language learning: an overview of sociocultural theory. In K. Van Esch & O. St.John (Eds.), *New insights into foreign language learning and teaching* (Vol. 9, pp. 13-34). Frankfurt am Main: Peter Lang GmbH.
Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
Lapkin, S. (1984). How well do immersion students speak and write French? *Canadian Modern Language Review, 40*(5), 576-585.
Larsen-Freeman, D. (2003). *Teaching language: From grammar to grammaring*. Boston: Thomson/Heinle.
Larsen-Freeman, D., & Long, M. (1991). *An Introduction to second language acquisition research*. London: Longman.
Lee, J. F. (1999). *Tasks and communicating in language classrooms*. New York: McGraw Hill.
Leont'ev, A. N. (1978). *Activity, consciousness and personality*. Englewood Cliffs, NJ: Prentice Hall.
Lewin, K. (1948). *Resolving social conflicts: Selected papers on group dynamics*. New York: Harper.
Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment* New york: Guilford.
Lightbrown, P., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: effects on second language acquisition. *Second Language Acquisition, 12*(429-448).
Lightbrown, P. M., & Spada, N. (1999). How languages are learned: Oxford: Oxford University Press.
Lin, M., & Mackay, C. (2004). Thinking through modern foreign languages. In D. Leat (Ed.). Cambridge: Chris Kington Publishing.
Little, D. (1991). Learner autonomy 1: Definitions, issues and problems. *Dublin: Authentik, 7*.
Little, D. (1994). Learner autonomy: A theoretical construct and its practical application. *Die Neueren Sprachen, 95*(5), 430-442.
Little, D. (1997). Language awareness and the autonomous language learner. *Language Awareness, 6*(n2&3), 93-104.
Little, D. (1999). *The European language portfolio and self-assessment*. Strasbourg: Council of Europe, document DECS/EDU/LANG (99)30.
Little, D., Ridley, J., & Ushioda, E. (2002). *Towards greater learner autonomy in the foreign language classroom*. Dublin: Authentik.
Long, M. (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition, 5*, 177-193.

Long, M. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly, 24*(4), 649-665.
Lowijck, J., & Verloop, N. (1995). *Onderwijskunde: Een kennisbasis voor professionals*. Groningen: Wolters-Noordhoff.
Luria, A. R. (1961). Study of the abnormal child. *American Journal of Orthopsychiatry. A Journal of Human Behavior, 31*, 1-16.
Luria, A. R. (1982). *Language and cognition*. New York: John Wiley and Sons.
Mackey, A., & Phlip, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings. *The Modern Language Journal, 82*, 338-356.
Mackey, W. F. (1965). *Language teaching analysis*. London: Longman.
Mahoney, M. (2004). What is constructivism and why is it growing? [Electronic Version]. Retrieved March 2006 from http://www.constructivism123.com/What_is/What_is_constructivism.html.
Marijnissen, J. (2003). *Interactief leren determineren: Een onderzoek naar de ontwikkeling en evaluatie van de interactieve flora van Nederland en Vlaanderen.* Radboud University, Nijmegen.
Markee, N. (2000). *Conversational analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
Maslow, A. H. (1970). *Motivation and personality*. New York: Harper and Row.
Mathijsen, I. C. H. (2006). Denken en handelen van docenten: Utrecht: IVLOS (diss.).
Matthijsen, M. A. J. M. (1972). *Klasse-onderwijs: Sociologie van het onderwijs*. Deventer: Van Loghum Slaterus.
Mc Collam Wiebe, K. (2002). *Learner readiness versus instructional type in L2 Spanish stage development*. Paper presented at the AAAL Conference Salt Lake City.
McCarthy, M., & Carter, R. (1994). *Language as discourse: Perspectives for Language Teaching*. New York: Longman.
McClelland, D. (1973). Testing for competence rather than for 'intelligence'. *American Psychologist, 28*, 1-14.
McCormick, D., & Donato, R. (2000). The discourse of teacher questions as scaffolding in an integrated ESL classroom. In J. k. Hall & L. S. Verplaetse (Eds.), *Second and foreign language learning through classroom interaction* (pp. 183–201). Philadelphia: Lawrence Erlbaum Associates.
McLaren, P. (1989). *Life in schools: An introduction to critical pedagogy in the foundations of education*. New York: Longman.
McNamara, T. (1996). *Measuring second language performance*. London: Longman.
McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
McNamara, T. (2002). *Insider, outsider: Language tests as weapons in intergroup conflict*. Paper presented at the Language Assessment Ethics Conference.
Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.
Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.
Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills: Sage Publications
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Beverly Hills: Sage Publications.
Minick, N. (1987). Implications of Vygotsky's theories for dynamic assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactive approach to evaluating learning potential*. New York: The Guilford Press.

Ministerie van Onderwijs Cultuur en Wetenschappen. (1998a). *Examenprogramma's profielen vwo/havo: Nederlands/moderne vreemde talen*. Den Haag: Centrale Financiële Instellingen (CfI).
Ministerie van Onderwijs Cultuur en Wetenschappen. (1998b). *Attainment targets 1998-2003: Basic secondary education*. The Hague: SDU.
Morrow, K. (1986). The evaluation of tests of communicative performance. *Innovations in language testing*, 29–42.
Morrow, K. (Ed.). (2004). *Insights from the Common European Framework*. Oxford: Oxford University Press.
Mulder, J. (1997). *Bronnenboek: Literatuur in het studiehuis*. Zutphen: Thieme Meulenhoff.
Musumeci, D. (1996). Teacher-learner negotiation in content-based instruction: Communication at cross-purposes? *Applied Linguistics, 17*(3), 286-325.
Musumeci, D. (2004). Research insights and communicative language teaching. In K. Van Esch & O. St.John (Eds.), *New insights into foreign language learning and teaching* (Vol. 9). Frankfurt am Main: Peter Lang Verlag.
Myers, I. (1962). *The Myers-Briggs type indicator*. Palo Alto, CA: Consulting Psychologists Press.
Nassaji, H., & Wells, G. (2000). What's the use of "triadic dialogue"?: An investigation of teacher-student interaction. *Applied Linguistics, 21*(3), 376-406.
Neisser, U. (1967). *Cognitive psychology*. New York: Appleton.
Newman, F., & Holzman, L. (1993). *Lev Vygotsky. Revolutionary scientist*. London: Routledge.
North, B. (2004). Relating assesments, examinations, and courses to the CEF. In K. Morrow (Ed.), *Insights from the Common European Framework* (pp. 77-90). Oxford: Oxford University Press.
Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman Group ltd.
Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Second Language Acquisition, 17*, 109-148.
Oxford, R. (1990). *Language learning strategies: What every teacher should know*. New York: Newbury House.
Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307.
Palmer, H. (1917). *The scientific study and teaching of languages*. New York: World Book Company.
Paulston, C. (1974). Linguistic and communicative competence. *TESOL Quarterly, 8*, 347-362.
Peña, E. D., & Gillam, R. B. (2000). Dynamic assessment of children referred for speech and language evaluations. In C. S. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications*. Amsterdam: Elsevier.
Pennycook, A. (1997). Cultural alternatives and autonomy. In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning*. New york: Longman.
Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. Mahwah NJ: Lawrence Erlbaum Associates.
Peters, V., & Wester, F. (1995). An Introduction to the principles and practice of our method of analysis. In H. J. C. Pieterse (Ed.), *Desmond Tutu's message: A qualitative analysis*. Kampen, the Netherlands: Kok.
Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition, 11*, 63-90.
Piepho, H.E. (1979). *Kommunikative Didaktik des Englishunterrichts*. Limburg: Frankonius Verlag.
Piepho, H. E. (1974). *Kommunikative Kompetenz als übergeordnetes Lernziel im Englischunterricht*. Dornburg-Frickhofen: Frankonius Verlag.

Pike, K. (1967). *Language in relation to a unified theory of the structure of human behavior*. The Hague: Mouton Publishers.
Poehner, M. E., & Lantolf, J. (2005). Dynamic assessment in the language classroom. *Language Teaching Research, 9*, 1-33.
Prawat, R. (2000). Dewey meets the "Mozart of psychology" in Moscow: The untold story. *American Education Research Journal, 37*, 663-696.
Raimes, A. (1983). Tradition and revolution in ESL teaching. *TESOL Quarterly, 17*, 535-552.
Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentagling the construct of formative assessment. *Language Testing, 17*, 215-243.
Reid, J. (1995). *Learning styles in the EFL/ESL classroom*. Boston: Heinle and Heinle.
Reis, S. (2005). *Learning to teach reading in English as a foreign language.* Thesis Radboud University, Nijmegen, the Netherlands.
Rivers, W. (1968). *Teaching foreign language skills* Chicago: The University of Chicago Press.
Rivers, W. (1973). From linguistic competence to communicative competence. *TESOL Quarterly, 7*, 25-34.
Rivers, W. (1975). *A practical guide to the teaching of French*. New York: Oxford University Press.
Rogers, C. R. (1983). *Freedom to learn for the Eighties*. Columbus OH: C. E. Merrill Publications.
Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3 ed., pp. 376-391). New York: MacMillan.
Rosenshine, B. V. (1986). Synthesis of research on explicit teaching. *Educational Leadership, 43*(7), 60-69.
Ross, M., & Fletcher, G. J. O. (1985). Attribution and social perception. In E. Aronson & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 73-122). Hillsdale, NJ: Lawrence Erlbaum Associates.
Rumelhart, D. E., & Ortony, A. (1977). Representation of knowledge. In R. C. Anderson, R. J. Spiro & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Savignon, S. J. (1972). *Communicative competence: An experiment in foreign language teaching*. Philadelphia: Philadelphia Center for Curriculum Development.
Savignon, S. J. (1983). *Communicative competence : Theory and classroom practice. Texts and contexts in second language learning*. Reading, Mass.: Addison-Wesley.
Savignon, S. J. (1997). *Communicative competence : Theory and classroom practice. Texts and contexts in second language learning* (2nd ed.). New York: McGraw-Hill.
Savignon, S. J. (2002). *Interpreting communicative language teaching : Contexts and concerns in teacher education*. New Haven: Yale University Press.
Schäffner, C., & Wenden, A. (1995). *Language and peace*. Brookfield, VT: Dartmouth Publishing Company.
Schalkwijk, E., van Esch, K., Elsen, A., & Setz, W. (2002). Learner autonomy and language teacher education. In S. J. Savignon (Ed.), *Interpreting communicative language teaching: contexts and concerns in teacher education* (pp. 165–190). New Haven: Yale University Press.
Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Schön, D. A., & Argyris, C. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisc: Jossey-Bass Publishers.
Schröder, K. (1980). Kleine Chronik zur Frühzeit des Fremdsprachenlernens und des Fremdsprachenunterrichts *Die Neueren Sprachen, 79*, 114-135.

Schröder, K., & Weller, F. (1975). Erwägungen zum Begriffsfeld Methodik aus fremdsprachendidaktischer Sicht *Die Neueren Sprachen, 24*, 210-233.
Schwab, J. J. (1978). *Science, curriculum, and liberal Education*. Chicago: Chicago University Press.
Sercu, L. (2000). *Acquiring intercultural communicative competence from textbooks. The case of Flemish adolescent pupils learning German*. Leuven: Leuven University Press.
Sercu, L. (2002). Autonomous learning and the acquisition of intercultural communicative competence. Some implications for course development. *Language, Culture and Curriculum, 15*(1), 61-74.
Sercu, L., & St.John, O. (2007). Teacher beliefs and their impact on teaching practice: A literature review In M. Jiménez-Raya & L. Sercu (Eds.), *Challenges in teacher development: Learner autonomy and intercultural competence*. Frankfurt am Main: Peter Lang GmbH.
Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics, 15*, 188-211.
Shohamy, E. (2001). *The Power of tests: A critical perspective on the uses of language tests*. Harlow: Pearson Education Ltd.
Skinner, B. F. (1957). *Verbal behavior*. New york: Appleton-Century-Crofts.
Skinner, B. F. (1971). Beyond freedom and dignity. New York: Knopf.
Smith, F. (1975). *Comprehension and learning: A conceptual framework for teachers*. New York: Holt, Rinehart and Winston.
Smits, T. (2003). *Werken aan kwaliteitsverbetering van leerlingonderzoek: Een studie naar de ontwikkeling en het resultaat van een scholing van docenten.* Thesis Radboud University, Nijmegen.
Sommers, N. (1982). Responding to student writing. *College, Composition, and Communication, 33*(2), 148-156.
Spada, N., & Lightbrown, P. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal, 83*, 1-22.
Spolsky, B. (Ed.). (1978). *Approaches to language testing*. Arlington, VA: Center for Applied Linguistics.
Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press
Spolsky, B. (1998). *What is the user's perspective in language testing?* Paper presented at the colloquium 'The State of the Art in Language Testing: The User's Perspective'.
Spolsky, B. (Ed.). (1999). *Concise encyclopedia of educational linguistics*. Oxford: Elsevier Science Ltd.
Spolsky, B. (1999). Language testing. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics* (pp. 695-703). Oxford: Elsevier Science Ltd.
Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing. The nature and measurement of learning potential*. Cambridge: Cambridge University Press.
Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Newbury Park Ca.: Sage Publications.
Strevens, P. (1972). Language teaching. In T. Sebeok (Ed.), *Current trends in linguistics: Linguistics in Western Europe* (Vol. 9, pp. 702-732). The Hague: Mouton.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development In S. Gass & C. Madden (Eds.), *Input in second language acquisition*. Rowley, MA: Newbury House.
Swain, M. (1988). Manipulating and complementing content teaching to maximize second language learning. *TESL Canada Journal, 6*, 68-83.
Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review, 50*(1), 158-164.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-98). London: The Continuum International Publishing Group.

Swain, M., & Lantolf, J. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In *Sociocultural theory and second language learning* (pp. 97-115). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal, 83*, 320-338.

Swan, M. (2007). *Practical English usage*. Oxford: Oxford University Press.

Taylor, C. S., & Nolen, S. B. (1996). What does the psychometrician's classroom look like?: Reframing assessment concepts in the context of learning [Electronic Version]. *Education Policy Analysis Archives*, *4*. Retrieved 2nd March 2007 from http://epaa.asu.edu/epaa/v4n17.html.

Taylor, D. (1994). Inauthentic authenticity or authentic inauthenticity? [Electronic Version]. *TESL-EJ*, *1*, 1-12. Retrieved 14th July 2006 from http://www.zait.uni-bremen.de/wwwgast/tesl_ej/ej02/a.1.html.

The Dutch Ministery of Education Culture and Science (O C & W). (1998). *Attainment targets 1998-2003: Basic secondary education in the Netherlands*. The Hague: SDU.

Theunissen, M. (1995). *Samen nascholen, maar hoe?* , Garant, Leuven-Apeldoorn.

Thorndike, E. L. (1932). *The fundamentals of learning*. New York: Columbia University

Thorne, S. L. (2004). Cultural historical activity theory and the object of innovation. In K. Van Esch & O. St.John (Eds.), *New insights into foreign language learning and teaching* (pp. 51-70). Frankfurt am Main: Peter Lang GmbH.

Titone, R. (1968). *Teaching foreign languages: An historical sketch*. Washington, DC: Georgetown University press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, Mass: Harvard University Press.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham: Open University Press.

Trim, J. L. M. (2001). *The work of the Council of Europe in the field of modern languages, 1957-2001*. Paper presented at the Symposium to mark the European Day of Languages in Graz 26th September 2001.

Trim, J. L. M., Coste, D., & North, B. (Eds.). (1998). *Modern Languages: Learning, Teaching, Assessment. A Common European Framework of Reference*. Strasbourg: Council of Europe.

Tweede Fase Adviespunt. (2005). *Zeven jaar tweede fase: Een balans*. Den Haag: Tweede Fase Adviespunt.

Valdman, A. (1980). Communicative ability and syllabus design for global foreign language courses. *Studies in Second Language Acquisition, 3*, 81-96.

Valette, R. (1977). *Modern language testing* (2nd ed.). New York: Harcourt Brace Jovanovich.

Valsiner, J., & van der Veer, R. (2000). *The social mind. Construction of the idea*. Cambridge: Cambridge University press.

Van Amelsvoort, J. (1999). *Perspectief op instructie, motivatie en zelfregulatie.* Radboud Universiteit, Nijmegen.

Van den Akker, J., & Bergen, T. C. M. (1997). Het leren door docenten. *VELON. Tijdschrift voor Lerarenopleiders 18*(4), 26-33.

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research*. London: Routledge.

Van Ek, J. A., & Alexander, L. G. (1975). *Threshold level English*. Oxford: Pergamon Press.

Van Els, T., Bongaerts, T., Extra, G., Van Os, C., & Janssen van Dieten, A. (1984). *Applied linguistics and the learning and teaching of foreign languages* (Rev. English ed.). Baltimore, Md: E. Arnold.

Van Esch, K., Schalkwijk, E., Elsen, A., & Setz, W. (1999). Autonomous learning in foreign language teacher training. In P. Faber, W. Gewehr, M. Jiménez-Raya & A. J. Peck (Eds.), *English teacher education in Europe: New trends and developments* (Vol. 1). Frankfurt an Main: Peter Lang Verlag.

Van Esch, K., & St.John, O. (2003). *A framework for freedom: Learner autonomy in foreign language teacher education* (Vol. 8). Frankfurt am Main: Peter Lang Verlag.

Van Lier, L. (1995). *Introducing language awareness*. London: Penguin Press.

Van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy, and authenticity*. London: Longman.

Van Lier, L. (2000). From input to affordances: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Van Lier, L. (2007). Action-based teaching, autonomy and identity. *Innovation in Language Learning and Teaching, 1*(1), 46-65.

Van Veen, K. (2003). *Teachers' emotions in a context of reforms.* Thesis Radboud University Nijmegen.

Van Vlasselaer. (1972). Le strucuro-global vis-à-vis des structuralismes. *Revue de Phonétique Appliquée, 21*, 25-30.

Vandergrift, L., & Bélanger, C. (1998). The national core French assessment project: Design and field test of formative evaluation instruments at the intermediate level *The Canadian Modern Language Review, 54*, 553-578.

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*(3), 287-301.

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features and research methodology. *AILA Review, 11*, 27-36.

VanPatten, B. (1996). *Input processing and grammar Instruction in second language acquisition*. Norwood, NJ: Ablex.

Viëtor, W. (1882). Der Sprachunterricht muss umkehren (reprinted 1965). In K. Flechsig (Ed.), *Neusprachlicher Unterricht*. Weinheim: Verlag Julius Beltz.

Villamil, O. S., & de Guerrrero, M. C. M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics, 19*(4), 491-514.

Villareal, J. J. (1947). Test of aural comprehension. *Speech Monographs, 15*, 121-132.

Vygotsky, L. S. (1956). *Isbrannye Psikhologicheskie Issledovaniya (Selected Psychological Investigations)*. Moscow: Izdatel'stvo Akademii Pedagogischeskikh Nauk SSSR.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (Ed.). (1987). *The collected works of L.S. Vygotsky, volume 1. Problems of general psychology. Including the volume Thinking and speech*. New York: Plenum Press.

Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language Testing, 10*(1), 41.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53*(4), 5-23.

Watson, J. B. (1913). Psychology as the behaviorist views It. *Psychological Review, 20*, 158-177.

Weiner, B. (1974). *Achievement motivation and attribution theory*. Morristown, NJ: General Learning Press.

Weiner, B. (1980). *An attributional theory of motivation and emotion*. New York: Springer Verlag.
Wenden, A. (1992). *Learner strategies for learner autonomy*. New York: Prentice Hall.
Wenden, A., & Rubin, J. (1987). *Learner strategies in language learning*. London: Prentice Hall.
Wertsch, J.V. (1998). *Mind as action*. Oxford: Oxford University Press.
Wesche, M., & Ready, D. (1985). Foreigner talk in the university classroom. In S. Gass & C. Madden (Eds.), *Input in second language acquisition*. Rowley, MA: Newbury house.
White, L. (1991). Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom. *Second Language Research, 7*(133-161).
Widdowson, H. G. (1990). *Aspects of language teaching.* Oxford: Oxford University Press.
Wilhelm, F. A. (2005). *English in the Netherlands: A history of foreign language teaching 1800-1920*. Utrecht: Gopher Publishers.
Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge: Cambridge University Press.
Wong Filmore, L. (1992). Learning a language from learners. In C. Kramsch & S. McConnell-Ginet (Eds.), *Text and context: Cross-disciplinary perspectives on language study* Lexington, MA: DC Heath.
Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry 17*, 89-100.
Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making, and classroom practice*. Cambridge: Cambridge University Press.
Yin, R. (1993). *Applications of case study research*. London: Sage Publications.
Yin, R. (1994). *Case study research: Design and methods* (2nd ed.). London: Sage Publications.
Young, D. J. (Ed.). (1999). *Affect in foreign language and second language learning*. New York: McGraw-Hill.
Zeichner, K. M., & Liston, D. P. (1996). *Reflective teaching: An introduction*. Mahwah NJ: Lawrence Erlbaum Associates.
Zhang, L. J. (2001). Awareness in reading: EFL students' metacognitive knowledge of reading strategies in an acquisition-poor environment. *Language Awareness, 10*(4), 268-288.
Zuengler, J. (1989). Performance variation in NS/NSS interactions: Ethnolinguistic difference or discourse domain? In S. Gass, C. Madden, D. Preston & L. Selinker (Eds.), *Variation in Second Language Acquisition: Discourse and Pragmatics*. Clevedon: Mutilingual Matters.
Zuengler, J., & Bent, B. (1991). Relative knowledge of content domain: An influence on native/non-native conversations. *Applied Linguistics, 12*, 397-415.

SAMENVATTING

De samenvatting van het proefschrift *Autonomie in de praktijk getoetst* bestaat uit twee delen. Allereerst worden de vragen gepresenteerd die in het onderzoek centraal staan. Vervolgens wordt een samenvatting gegeven van de inhoud van de twaalf hoofdstukken, met aan het einde de aanbevelingen voor vervolgonderzoek en voor cursussen die zijn gericht op positieve terugslageffecten van toetsen op hoe docenten onderwijzen en hoe leerlingen leren.

Onderzoeksvragen

Het onderzoek bentwoordt drie hoofdvragen, die in de loop van het onderzoek zijn uitgewerkt in zeven meer specifieke vragen.

1. Wat vinden eerstegraads docenten Engels van actief en zelfstandig leren, communicatief taalonderwijs en de rol van beoordeling en evaluatie?

- Wat kan men leren van het bestuderen van theorie over de *autonomie van de leerling*, *communicatief taalonderwijs* en *beoordeling en evaluatie* van kennis van en vaardigheden in een vreemde taal?
- Hoe interpreteren en definiëren eerstegraads docenten Engels bovenstaande drie begrippen?
- Welke overtuigingen, ervaringen en argumenten liggen ten grondslag aan het ontwerp en de afname van drie informele voorbeeldtoetsen die door de docenten zelf zijn gemaakt?

2. In hoeverre vindt men de *beliefs* van drie informanten terug in hun beoordelings- en evaluatiepraktijk?

- Hoe interpreteren de docenten de toetsresultaten in relatie tot hun *beliefs* over communicatief taalonderwijs en de autonomie van de leerling?
- In hoeverre komen de resultaten van dit onderzoek overeen met de theoretische verkenningen van de drie centrale begrippen in deze studie?

3. Welke aanbevelingen kunnen worden gedaan voor het stimuleren van leeromgevingen waarin toetsen een positief terugslageffect hebben op de wijzen waarop leerlingen leren communiceren in het Engels?

- In hoeverre hebben de vernieuwingen van de Tweede Fase bijgedragen aan het vergroten van de autonomie van de leerlingen, het stimuleren van hun communicatieve vaardigheden in het Engels en het ontwikkelen van een effectieve beoordelings- en evaluatiepraktijk?
- Wat is de opbrengst voor vervolgonderzoek en cursussen gericht op positieve terugslageffecten van toetsen op hoe docenten onderwijzen en hoe leerlingen leren?

De hoofdstukken

Het **eerste** hoofdstuk, *Testing for autonomy: An introduction*, begint met twee persoonlijke gebeurtenissen waardoor de onderzoeker geïnteresseerd is geraakt in het belang van heldere toetscriteria en het effect dat een toets kan hebben als leerlingen een docent vragen ‘of het voor een punt is’. Vervolgens wordt uitgelegd dat de Engelse titel van de studie, *Testing for autonomy,* drie betekenissen heeft. De

titel verwijst allereerst naar de grenzen van autonomie die de onderzoeker samen met zijn collega's opzoekt in zijn werk als lerarenopleider. De titel verwijst tevens naar de manier waarop eerstegraads docenten Engels de grenzen van autonomie opzoeken bij leerlingen in de vierde klas van het havo en vwo. Tot slot verwijst de titel naar de wisselwerking tussen een theoretische verkenning van het construct en het verzamelen van data uit de lespraktijk. Hierna wordt ingegaan op de belangrijke rol die zowel overtuigingen (*beliefs*) als de specifieke sociaal-culturele context spelen in de alledaagse praktijk van docenten.

In het **tweede** hoofdstuk, The exploratory multiple case study, worden de relevantie van het onderzoek en de gemaakte methodologische keuzes besproken. Daarna wordt achtereenvolgens ingegaan op de doelen, vragen en gehanteerde methodiek van het onderzoek. De keuze voor een explorerende meervoudige case study wordt verantwoord. Tevens wordt ingegaan op de notie *analytical generalizability* (Yin, 1994: 36), waaruit blijkt dat gevalstudies niet representatief zijn en dat de resultaten alleen kunnen worden gegeneraliseerd naar theorie. In het proces van data-analyse kunnen nieuwe theorieën ontstaan, bestaande theorieën worden ontkracht of juist bevestigd. In het hoofdstuk komt tevens aan bod hoe de drie *good practice* docenten zijn geselecteerd. Voor gevalstudies is het van belang dat onderzoek systematisch en nauwgezet wordt geprotocolleerd en gedocumenteerd. Het hoofdstuk geeft daarom ook inzicht in de vijf stadia van dataverzameling.

Het **derde** hoofdstuk, *Learner autonomy as a pedogical construct*, is een theoretische verkenning van het construct *autonomie van de leerling*. In de verkenning wordt besproken wat autonomie is, waarom autonomie kan worden gezien als een haalbaar pedagogisch doel en wat de relatie is tussen motivatie en autonomie wanneer leerlingen leren communiceren in een andere taal.

De vraag wat autonomie is, wordt beantwoord door gangbare werkdefinities van het begrip te bespreken. Deze bespreking levert dertien parameters op van autonomie. De onderscheiden parameters zijn achtereenvolgens:

- de mate waarin zowel docent als leerling worden *geëngageerd en gemotiveerd* om autonomie als pedagogisch doel na te streven;
- de ontwikkeling en het in stand houden van een *krachtige en uitdagende leeromgeving*;
- de specifieke *kennis, vaardigheden en motivatie* van leerlingen om effectief te leren;
- de specifieke *leerdoelen van het curriculum*;
- de geselecteerde *inhoud* van het curriculum en het *tempo* waarmee de inhoud wordt verwerkt;
- de specifieke *taken, handelingen en activiteiten* die leerlingen uitvoeren;
- de *leerstrategieën en technieken* die daarvoor nodig zijn;
- de mate van *transfer* van leertaken, handelingen en activiteiten;
- de mate waarin leerlingen *zich verantwoordelijk voelen* voor hun eigen leerproces;
- de mate waarin leerlingen *zich concentreren, doorzetten* en *moeite doen*;
- de mogelijkheden voor leerlingen om te *zeggen wat zij voelen en denken*;
- de wijzen waarop voortgang wordt *gevolgd, beoordeeld* en *geëvalueerd*;
- de rollen van *metacognitieve* en *metalinguistische reflectie*.

Bij de bespreking van de parameters wordt benadrukt dat autonomie een buitengewoon complex en gelaagd begrip is, dat niet in het regulier voortgezet onderwijs kan worden geïmplementeerd zonder een docent die veel weet en veel

kan. Het is een misvatting dat een leerling het zonder een docent kan stellen in de groei naar meer kennis, vaardigheden en autonomie.

De vraag waarom de autonomie van de leerling als een haalbaar pedagogisch doel moet worden nagestreefd, wordt beantwoord in een multidisciplinaire verkenning van het begrip.

Een sociaal-historische verkenning van autonomie levert drie argumenten op. Allereerst is het nastreven van autonomie geen nieuw pedagogisch doel. Zowel Socrates als Comenius, docenten uit het verleden met enige reputatie, streefden de autonomie na van de leerlingen aan wie zij les gaven. Vervolgens wordt besproken dat een democratische cultuur-oriëntatie als enige van de vier besproken cultuuroriëntaties (Matthijssen, 1972) een model van zelfbepaling en autonomie met zich meebrengt. Een derde argument is dat onze kennismaatschappij in toenemende mate om mensen vraagt die in staat zijn actief en zelfstandig te leren.

Een filosofische verkenning van het begrip autonomie levert een vierde rechtvaardiging op voor autonomie als pedagogisch doel. Binnen de filosofie bestaan grofweg vier benaderingswijzen en interpretaties van autonomie, die elkaar deels overlappen (Buss, 2002). In *coherentist* interpretaties van het begrip, zijn mensen autonoom als zij de motieven die hen aanzetten tot bepaald denken of handelen accepteren, rechtvaardigen, zich ermee identificeren of geloven dat zij betekenisvol zijn in relatie tot hun plannen of verplichtingen. In *reasons-responsive* benaderingswijzen van autonomie wordt beargumenteerd dat mensen alleen autonoom zijn als hun denken of handelen is gebaseerd op een breed scala van argumenten, die hen al dan niet aanzet tot actie. Aanhangers van *responsiveness-to-reasoning* interpretaties van autonomie benadrukken dat mensen alleen autonoom zijn als zij in staat zijn de motieven te evalueren die hen aanzetten tot bepaald denken of handelen. In die evaluatie worden de oorspronkelijke beweegredenen tot handelen waar nodig bijgesteld. Tot slot zijn er filosofen die vinden dat persoonlijke autonomie niet bestaat. Zij zijn die mening toegedaan, omdat de motieven voor denken en handelen die ontstaan uit de persoon zelf niet op een valide manier kunnen worden onderscheiden van motieven die buiten de persoon ontstaan. Deze benaderingswijze wordt *incompatibilist* genoemd. Met name de *reasons-responsive* en de *responsive-to-reasoning* benaderingswijzen van het menselijk denken en doen bieden concrete ontwikkelingsperspectieven voor het begrip autonomie in onderwijsleersituaties: de leerlingen denken na over en ontwikkelen een breed scala aan argumenten die hen aanzetten tot bepaald denken en handelen, en kunnen deze evalueren in relatie tot andere argumenten.

Een vijfde rechtvaardiging voor autonomie als pedagogisch doel komt voort uit een politieke benadering van het begrip autonomie. Aanhangers van een politieke interpretatie van autonomie wijzen op de noodzaak de machtsstructuren te herzien die gewoonlijk bestaan tussen docent en leerling en te komen tot meer zeggenschap en invloed van de leerling op het leren en onderwijzen. Toenemende zeggenschap en invloed van de leerling op het onderwijsleerproces kan een belangrijke factor zijn in de ontwikkeling van meer autonomie bij de leerlingen.

Het zesde en laatste argument voor autonomie als pedagogisch doel komt voort uit vier verschillende oriëntaties op leren binnen de leerpsychologie. De vier benaderingen streven naar het vergroten en vermeerderen van de autonomie van de leerling. De besproken oriëntaties komen voort uit de cognitieve psychologie (Ausubel 1963, 1968), de humanistische psychologie (Maslov, 1970; Rogers, 1983), het constructivisme (Mahoney, 2004) en de sociaal-culturele theorie (Vygotsky, 1956, 1978, 1987; Lantolf & Thorne, 2006).

Aangezien het motiveren en de motivatie van leerlingen een rol spelen binnen de besproken leertheorieën, wordt ook de relatie tussen autonomie en motivatie nader besproken. Het hoofdstuk benadrukt de relevantie van het onderscheid tussen intrinsieke en extrinsieke motivatie (Deci, 1975; Deci & Ryan, 1985) en integratieve en instrumentele motivatie (Gardner & Lambert, 1972; Gardner & McIntyre, 1991).

De attributietheorie (Heider, 1958; Weiner, 1994, 1980; Ross & Fletcher, 1985) wijst op het belang dat leerlingen leren hun leersucces toe te schrijven aan zaken waarover zij zelf controle hebben. Het motivationele model van DeCharms (1984) toont het belang dat leerlingen leren hun eigen leren te initiëren. Tot slot wordt het pleidooi van Dickinson (1995) en Van Lier (1996) ondersteund voor meer aandacht voor intrinsieke motivatie in het reguliere onderwijs en voor het belang van twee belangrijke variabelen als leerlingen leren leren: controle over het leren en de sociaal-culturele context waarbinnen het leren plaatsvindt.

Het **vierde** hoofdstuk, *Communicative competence in foreign language education*, richt zich op de achtergronden en inhoud van het begrip communicatieve competentie bij het leren communiceren in een andere taal. Het hoofdstuk begint met een overzicht van methodologische benaderingswijzen voor het leren en onderwijzen van andere talen. Achtereenvolgens worden besproken de grammatica-vertaalmethode, de directe methode, de reform-methode, de audiolinguale methode en de audiovisuele methode. Tot slot worden de meer audiolinguaal gerichte methoden vergeleken met meer communicatief gerichte methoden. De conclusie is dat in discussies over methodologische benaderingswijzen in het huidige communicatieve tijdperk noch het verleden moet worden vergeten, noch de toekomst kritiekloos tegemoet moet worden gezien.

Vervolgens wordt het begrip communicatieve competentie nader besproken en gedefiniëerd. Het baanbrekende werk van de taalkundige en antropoloog Hymes (1967,1971, 1974) wordt genoemd. Zijn visie op taal als sociaal gedrag opende de weg voor een sociaal-interactionistische benadering van communicatie. Voor communicatie blijkt meer noodzakelijk dan woorden en grammatica. Halliday (1970,1973,1978) heeft zich verdiept in de specifieke functies van taal en daarmee in belangrijke mate bijgedragen aan interpretaties van communicatieve competentie.

De aandacht voor communicatieve competentie leidt uiteindelijk tot een model dat nog steeds relevant is en van waaruit aanvullende modellen zijn ontwikkeld. Het betreft het model communicatieve competentie van Canale & Swain (1980), dat later is verfijnd door Canale (1983). Canale & Swain definiëren communicatieve competentie als een samenhangend geheel van vier deelcompetenties: grammaticale competentie, discursieve competentie, sociolinguistische competentie en strategische competentie. Zo ontstond een definitie van communicatie (Savignon, 1983, 1997) als “een voortdurend proces van het uiten, interpreteren en bespreken van betekenis” (1997:14). Het is een definitie die verder gaat dan leerlingen grammaticaregels en woordenlijsten laten reproduceren en toepassen.

Vanuit een meer didactisch perspectief gaat het hoofdstuk vervolgens in op drie verschillende benaderingswijzen bij het aanleren van communicatieve competentie: *van vorm naar betekenis* (Paulston, 1974; Rivers, 1972, Valette, 1977), van *betekenis naar vorm* (Piepho, 1974, 1979; Widdowson, 1990) en *specificatie van de context* (Council of Europe, 1998, 2001). Naast diversiteit is ook sprake van gemeenschappelijkheid in communicatieve benaderingswijzen. (Berns, 1990) noemt acht aspecten. Allereerst is het *leren communiceren* het hoofddoel van communicatief taalonderwijs. Taal wordt gezien als een sociaal instrument dat betekenis verleent en helpt om betekenis te verkrijgen. Gebruikers communiceren mondeling of schriftelijk met iemand over iets met een bepaald doel. Een tweede aspect is de erkenning van *diversiteit* bij taalontwikkeling en taalgebruik als leerlingen leren communiceren in een andere taal. Een derde aspect is dat de communicatieve vaardigheid van een leerling in relatieve en niet in absolute termen van *correctheid* wordt gezien. Een vierde aspect is de erkenning en acceptatie van meerdere *taalvarianten* bij het leren en onderwijzen van een andere taal. Ten vijfde wordt een bepalende rol toegekend aan *cultuur*, zowel bij het leren en verwerven van de moedertaal als bij andere talen. Een zesde gemeenschappelijk aspect is dat geen enkele *methodologie* of vaststaande didactische procedures worden

voorgeschreven. Een zevende aspect is erkenning van de relatie tussen de te bereiken taalvaardigheid en de specifieke *functies* van taal, zoals het genereren van ideeën, het bevorderen van interpersoonlijk contact of het begrijpen van gesproken en geschreven teksten. Tot slot noemt Berns als gemeenschappelijk aspect dat leerlingen een taal leren door deze te *gebruiken* vanuit uiteenlopende doelen in alle fases van het leren. Daarbij wordt in toenemende mate het belang erkend van wat leerlingen ervaren, wensen of vinden (1990:104).

Het hoofdstuk eindigt met een overzicht van recent onderzoek op het gebied van communicatieve competentie, dat in grote lijnen is onderverdeeld in aandacht voor *betekenis*, *vorm* en *de culturele aspecten van taalgebruik*. Een belangrijke conclusie is dat de onderzoeksresultaten aanleiding geven voor gelijkwaardige aandacht voor zowel vorm, betekenis en cultuur als een leerling leert communiceren in een andere taal.

Het **vijfde** hoofdstuk, *Foreign language assessment en evaluation*, richt zich op het derde construct dat centraal staat in het dissertatie-onderzoek. Allereerst wordt ingegaan op het belang van het onderscheiden van de begrippen toetsing, beoordeling en evaluatie. Met name de begrippen beoordeling en evaluatie worden in discussies over toetsing met elkaar verward, zowel in de vakliteratuur als in de toetsingspraktijk. Beoordeling wordt gedefinieerd als de classificering van de kennis en/of vaardigheden van een getoetste op een bepaald moment in de tijd door middel van een toets of andere beoordelingsprocedure. Evaluatie wordt gezien als een retrospectieve en prospectieve procedure, waarin de resultaten van een toets of beoordeling worden geïnterpreteerd door zowel de toetser als de getoetste.

Vervolgens worden vier ontwikkelingen of trends besproken op het gebied van taaltoetsing. Deze zijn achtereenvolgens de *voorwetenschappelijke*, *psychometrisch-structuralistische*, *integratief-sociolinguistische* en de *kritisch-dynamische* trend. De genoemde trends, die min of meer naast elkaar bestaan, zijn geliëerd aan de eerder besproken visies op leren en interpretaties van communicatieve competentie.

Een bijzondere plaats is weggelegd voor de bespreking van de kritisch-dynamische trend. De trend is verwant met het postmodernisme van filosofen als Lyotard, Derrida, Foucault en Rorty, die de zekerheden in twijfel trekken die sinds de Verlichting zijn verkondigd door rationalisten en modernisten. Het hoofdstuk bespreekt een aantal kritisch-dynamische aandachtgebieden: *critical language testing*, collaboratieve en dialogische benaderingen, de maatschappelijke en ethische verantwoordelijkheid van taaltoetsers, de rechten van getoetsten, dynamische wijzen van beoordeling en evaluatie en de relatie tussen *dynamic assessment* en formatieve beoordeling en evaluatie.

Vervolgens gaat het hoofdstuk in op professionele standaarden op het gebied van taaltoetsing, te weten de relatie tussen toetsdoel en toetssoort en de essentiële toetskwaliteiten van betrouwbaarheid en validiteit. Er wordt in het bijzonder aandacht besteed aan het begrip constructvaliditeit (Messick, 1989) en mogelijke toepassingen in het reguliere voortgezet onderwijs (Taylor & Nolen, 1996).

Tot slot wordt het model *test usefulness* (Bachman & Palmer, 1996) besproken. In dit model zijn de eerder besproken professionele standaarden op het gebied van toetsing geïntegreerd. Bachman en Palmer noemen een toetstaak effectief als deze voldoet aan de kwaliteitseisen van *betrouwbaarheid, constructvaliditeit, authenticiteit, interactiviteit, positieve terugslageffecten* en *praktische haalbaarheid.* Het model helpt om de kwaliteit te vergroten van de voorbereiding, afname, beoordeling en evaluatie van overhoringen, toetsen en schoolexamens.

Het **zesde** hoofdstuk, *A context of innovation*, bespreekt de achtergrond van het onderwijs in Nederland op het moment van dataverzameling. Na het bespreken van de wettelijke parameter van vrijheid van onderwijs, wordt een overzicht gegeven van het onderwijs in Nederland en de opleiding tot leraar basisonderwijs en de

opleidingen tot respectievelijk tweede- en eerstegraadsdocent in het voortgezet onderwijs. Daarna wordt het vak Engels besproken in het basisonderwijs, de Basisvorming en de Tweede Fase. De bespreking richt zich op de leerdoelen en de wijzen waarop het bereiken van de gestelde leerdoelen wordt geëvalueerd. Een grootschalige peiling van de onderwijskwaliteit in de Basisvorming (Inspectie van het Onderwijs 1999a; 1999b) laat zien dat het moeilijk is de communicatieve doelen te behalen die voor het vak Engels en de overige moderne vreemde talen zijn vastgesteld.

Een beschrijving van de context van vernieuwing van de bovenbouw toont aan dat schoolleiders en docenten voor een zware taak zijn gesteld. Naast de invoering van een fundamenteel gewijzigd curriculum en veelal het terugbrengen van het aantal contacturen tussen docent en klas, wordt tevens van docenten verwacht dat zij het actief en zelfstandig leren van hun leerlingen bevorderen.

In het hoofdstuk wordt verder besproken dat de vernieuwingen top-down zijn geïntroduceerd, met weinig aandacht voor wat er op de werkvloer leeft. Zes aspecten ontbreken bij de officiële start van de vernieuwde Tweede Fase in 1999. Allereerst ontbreken betrouwbare, valide of aannemelijke onderzoeksgegevens over de implementatie van actief en zelfstandig leren gebaseerd op data uit havo en vwo-klassen. Een tweede aspect is het gebrek aan aandacht voor hoe de vernieuwingen van het curriculum en de didactiek worden ontvangen door de docenten, die geacht worden deze effectief te implementeren. Ten derde is er geen samenhang tussen specifieke vakconcepten, vakdidactiek en algemene didactiek, waardoor het actief en zelfstandig leren vooral algemeen didactisch lijkt te worden geïnterpreteerd. Een vierde aspect is de niet heldere plaatsbepaling van beoordeling en evaluatie in de Tweede Fase. De programma's van toetsing en afsluiting (PTA's) zijn grotendeels gericht op verkaveling van leerstof, beoordeling en het specifieke beoordelingsmoment. Als vijfde aspect wordt genoemd het grotendeels ontbreken van vakspecifieke voorbeelden van *good practice* op het moment van invoering van de Tweede Fase. Tot slot wordt genoemd het ontbreken van een fundamentele discussie over de vraag in hoeverre domeinspecifieke en vakspecifieke constructen zich verhouden tot de gewenste vernieuwingen van het curriculum en de didactiek.

Het **zevende** hoofdstuk, *Three stories to tell*, is het eerste datahoofdstuk van de studie. Het is een verhalend hoofdstuk dat is gebaseerd op de interviews met de drie infomanten voor de aanvang van het schooljaar waarin de data zijn verzameld. In het hoofdstuk zijn *Joy*, *the budding professional*, *Mark*, *the literary master* en *Pete, the project man* voor een belangrijk deel zelf aan het woord.

Het hoofdstuk heeft de volgende opbouw. Eerst worden de drie respondenten voorgesteld. Daarna vertellen de docenten over formatieve ervaringen in hun loopbaan, hun kernovertuigingen (*core beliefs*) en over hun interpretaties van autonomie van de leerling, communicatief taalonderwijs en effectieve toetsing.

De bespreking van de visies op effectieve toetsing valt uiteen in wat de respondenten zien als *essentiële kennis, vaardigheden en inzicht*, wat zij verstaan onder *een effectieve schriftelijke taaltoets Engels* en vervolgens wat zij zien als *terugslageffecten* van toetsen. Vervolgens vermelden de respondenten eventuele plannen of voornemens voor het eerste jaar waarin de Tweede Fase officieel wordt ingevoerd. Tot slot wordt verwezen naar de schriftelijke taaltoetsen die in de loop van het schooljaar door de respondenten zijn gekozen en met hen zijn besproken. De data van *Joy*, *Mark* en *Pete* uit het eerste interview worden hier achtereenvolgens gepresenteerd.

Joy noemt drie formatieve ervaringen die van belang zijn geweest in haar onderwijsloopbaan en toetsingspraktijk. De eerste formatieve ervaring doet zij op als studente wanneer zij een neefje bijles geeft. Door hem bepaalde aspecten van het Engels uit te leggen, verkrijgt zij zelf inzicht in het Engels en in het leren en

onderwijzen daarvan. De tweede formatieve ervaring doet zij op in de tweede school waar zij als leraar komt te werken. Hier leert zij het belang en nut van het kritisch evalueren van een leergang en daarvan af te wijken als zij dat noodzakelijk acht. Als derde formatieve ervaring leert zij, met name op haar tweede school, collegiale samenwerking waarderen en ziet deze samenwerking als meerwaarde in haar ontwikkeling als docent. Bovenstaande formatieve ervaringen leiden tot de volgende kernovertuigingen, oftewel *core beliefs*.

Joy gelooft dat:
- de gevorderde student of beginnende docent Engels inzicht ontwikkelt in hoe de taal werkt door zelf les te geven;
- het lezen en analyseren van literaire teksten van meerwaarde is voor het onderwijs Engels;
- inzicht in hoe een taal werkt een steeds grotere rol moet krijgen in haar onderwijs- en toetsingspraktijk;
- leraren hun eigen lesmateriaal en toetsen dienen te ontwikkelen, die zijn gebaseerd op authentieke teksten over thema's die haar leerlingen aanspreken en die betekenisvol zijn;
- leraren onderling nauw moeten samenwerken, bijvoorbeeld bij wat te toetsen op welk moment en bij het uitwisselen van ideeën en ervaringen;
- toetsen meer moeten beoordelen dan de korte-termijn reproductie van op zichzelf staande items die uit het hoofd zijn geleerd;
- drills en de reproductie van standaardzinnen leerlingen niet helpen leren communiceren in het Engels;
- te veel aandacht wordt besteed aan grammaticale zaken zoals het lijdend maken van bedrijvende zinnen; deze tijd en energie is beter besteed aan het leren gebruiken van de juiste grammaticale tijd of vorm in het Engels;
- het toetsen van idioom met behulp van authentieke Engelse teksten en schrijftoetsen nuttig is;
- woordverwerving gericht moet zijn op lange-termijn retentie, wat kan worden bereikt door leerlingen hun eigen idioombestanden aan te laten leggen en de leerlingen extensief authentieke teksten te laten lezen die iets boven hun niveau liggen;
- integratieve toetsen nuttiger en zinvoller zijn dan discrete-point toetsen;
- het belangrijk is leerlingen gerust te stellen als zij in paniek raken;
- het moeilijk is om schrijftoetsen betrouwbaar en valide te beoordelen;
- het leerlingen motiveert als zij zelf bepaalde keuzes mogen maken, vooral op terreinen waarvoor zij in eerste instantie niet zijn gemotiveerd, zoals bijvoorbeeld literaire teksten.

Joy heeft een duidelijke visie op de autonomie van de leerling. Zij gelooft dat de zelfstandigheid van leerlingen toeneemt als docenten hun leerlingen explicieter laten nadenken over wat zij aan het doen zijn, wat zij moeten doen, wat zij willen doen en wat er vervolgens nodig is om dat te bereiken. Zij voegt daaraan toe dat de groei naar meer autonomie geleidelijk moet zijn, omdat leerlingen gewend zijn taken uit te voeren zonder daar veel bij na te denken. Ook noemt zij dat volledig zelfverantwoordelijk leren in het voortgezet onderwijs moeilijk te realiseren is. Zowel de eisen en doelen van het voortgezet onderwijs als de persoonlijkheidseigenschappen van de leerlingen stellen grenzen aan de mate van zelfverantwoordelijkheid. De leraar zelf heeft een sturende rol in de ontwikkeling naar meer zelfstandigheid van de leerling.

Joy vindt het moeilijk te omschrijven wat zij verstaat onder communicatief taalonderwijs. Zij geeft aan wat zij niet accepteert en dat is een "alles-wat-je–zegt-of-schrijft- is-goed" aanpak. Zij wijst daarbij op het belang van essentiële kennis of

vaardigheden. De materialen die *Joy* gebruikt en de voorbeelden die zij noemt zijn authentiek en betekenisvol, hoewel zij hiernaar niet expliciet verwijst in haar definitie.

Een effectieve schriftelijke toets Engels wordt door *Joy* gedefinieerd als N+1, wat zij interpreteert als iets moeilijker dan wat normaal gesproken van een leerling mag worden verwacht. Zij benadrukt dat de geleerde kennis en/of vaardigheden bij voorkeur worden toegepast in authentieke en betekenisvolle toetstaken.

Als essentiële kennis, inzicht en vaardigheden in het Engels noemt zij vier aspecten. Allereerst is dat de vaardigheid om vragen en ontkenningen in het Engels te maken “op redelijk correcte wijze” (1-1-11). Dan noemt zij de vaardigheid om werkwoorden in de juiste grammaticale tijd en vorm te zetten. Vervolgens hecht zij belang aan expliciete kennis van de vormen van de onregelmatige werkwoorden in het Engels. Als laatste noemt zij het gebruik van de juiste woordvolgorde in het gesproken en geschreven Engels.

Met het oog op de terugslageffecten van toetsen, vindt *Joy* dat toetsen niet de instrumenten zijn die leerlingen leren communiceren in het Engels. Die rol is volgens haar weggelegd voor motiverende oefentaken. Hoewel zij erkent dat cijfers belangrijk zijn voor leerlingen, noemt zij de visie dat leerlingen alleen maar voor een cijfer leren een misvatting.

Mark

Mark noemt twee formatieve ervaringen die van belang zijn geweest in zijn ontwikkeling als eerstegraads docent. Allereerst verwijst hij naar zijn ervaringen als student met Engelse en Amerikaanse literatuur tijdens zijn universitaire studie. De aangeboden korte verhalen en romans en de wijzen waarop literaire teksten destijds werden besproken en getoetst, hebben zijn onderwijs- en toetsingspraktijk beïnvloed. *Mark* biedt zijn leerlingen een selectie aan van het corpus dat hij ooit zelf heeft bestudeerd, waarbij een literair begrippenkader wordt aangebracht dat leerlingen moet helpen de literaire werken te doorgronden en waarderen, met het essay, schriftelijk verslag en mondelinge tentamen als populaire toetsvormen. *Mark* verwijst naar een tweede formatieve ervaring. In zijn eerste jaren als docent op een gymnasium, geeft hij hoofdzakelijk les in de onderbouw. *Mark* vindt dat die periode bij uitstek geschikt is geweest om een grondig begrip van de Engelse grammatica te krijgen en te leren hoe deze kennis zo effectief mogelijk kan worden onderwezen. De twee genoemde formatieve ervaringen komen terug in de kernovertuigingen die hij noemt. *Mark* gelooft in:

- literatuuronderwijs en het analyseren van literaire teksten in het voortgezet onderwijs: door literatuur te lezen leer je over het leven;
- ex cathedra directe instructie, waarin een docent de leerlingen disciplineert en aanmoedigt tot prestaties;
- leraren die een voorbeeld zijn voor hun leerlingen in gedrag en in de wijzen waarop bepaalde literaire of grammaticale vakinhoud moet worden verwerkt;
- het transfereren van academische literaire kennis en vaardigheden naar het voortgezet onderwijs;
- een duidelijk onderscheid tussen de onder- en bovenbouw in het voortgezet onderwijs, waarbij in de onderbouw het accent wordt gelegd op het verkrijgen van een stevig fundament van grammaticale en idiomatische kennis, dat vervolgens in de bovenbouw wordt uitgebreid en toegepast op het lezen, bespreken en analyseren van literaire teksten;
- expliciet grammatica- en idioomonderwijs op een zo hoog mogelijk niveau, regelmatig getoetst in vertaalzinnen van het Nederlands naar het Engels;
- het geleidelijk opbouwen van literaire en linguistische kennis en vaardigheden door de schooljaren heen;
- het belang van contacturen tussen docent en klas en de noodzaak dat de leraar de leerlingen bij het leren begeleidt;

- de vrijheid en autonomie van leraren in hun denken en handelen. Die autonomie wordt aangetast door voorgeschreven vernieuwingen als de zinloos egalitaire Basisvorming en de komst van een rigide Tweede Fase;
- de stelling dat de rol van literatuur is gemarginaliseerd in het curriculum van de vernieuwde Tweede Fase;
- de ambivalentie van het gebruik van leergangen die enerzijds nodig zijn om het onderwijs vorm te geven en anderzijds zelden recht doen aan het niveau van gymnasiasten;
- de tekortkomingen van de toetsen die onderdeel uit maken van de gebruikte leergangen, zodat ze vaak door de leraar moeten worden aangepast;
- expliciete aandacht voor foutencorrectie en de leerlingen de fouten te laten analyseren die zij hebben gemaakt;
- het beoordelen van de kennis en vaardigheden aan het einde van het leerproces en niet tijdens het proces van het leren communiceren zelf;
- de positieve en noodzakelijke prikkels van toetsen, die leiden tot actie en gerichte prestaties;
- de opvatting dat gedwongen afname van de Basisvormingstoetsen Engels, die ver onder het niveau liggen van wat zijn leerlingen beheersen onzin is;
- de stelling dat het eindcijfer voor Engels dat een leerling behaalt onevenredig zwaar wordt bepaald door de resultaten op de nationale CITO leesvaardigheidstoets.

De autonomie van de leerling bestaat volgens *Mark* uit drie deelcompetenties: planningsvaardigheden, zelfbeoordelingsvaardigheden en de vaardigheid om denk- en doe-activiteiten beredeneerd met elkaar af te wisselen. Ook *Mark* benadrukt de sturende rol die de leraar speelt in de groei naar meer zelfstandigheid en zelfverantwoordelijkheid van zijn leerlingen. Hij schept de randvoorwaarden, stelt de eisen waaraan de leerlingen moeten voldoen en stuurt bij waar nodig.

De essentie van communicatief taalonderwijs is voor *Mark* dat communicatie een vaardigheid is voor het leven. Leraren leren leerlingen communiceren in het Engels door voorbeelden te geven, te instrueren en oefentaken aan te bieden die het mogelijk maken dat leerlingen van elkaar leren. *Mark* ziet grammaticaonderwijs en taalbeschouwing als belangrijke elementen van het leren communiceren in het Engels. Als voorbeeld noemt hij de semantische verschillen die leerlingen dienen te ontdekken tussen ‘This is John’s picture’; ‘This is a picture from John’; ‘This is a picture of John’s’; ‘This is a picture of John’.

Mark vindt dat een effectieve schrifteljke toets Engels zich moet verhouden tot het niveau dat een leerling op het toetsmoment dient te hebben bereikt. Verder wordt een effectieve toets gekenmerkt door toetstaken waarin de leerling op creatieve wijze de bestudeerde grammatica en het geleerde idioom moet toepassen.

Van essentiële kennis, inzicht en vaardigheden in het Engels noemt *Mark* vier elementen. Allereerst merkt hij op dat de essentiële kennis en vaardigheden voor een groot deel worden bepaald door de leergangen die op school worden gebruikt. Hij noemt tevens het belang van de juiste spelling en een correcte wijze van formuleren, die voor hem indicatoren zijn van effectieve stijl. Vervolgens noemt hij literair inzicht, met in het bijzonder de vaardigheid om eigen meningen over de inhoud van literaire teksten in het Engels te formuleren. Tot slot hecht *Mark* belang aan grammaticaal inzicht, dat hij interpreteert als kennis van de grammaticale en semantische relaties tussen de woorden en zinsdelen in een zin.

Mark noemt drie terugslageffecten van toetsen. Alleerst is hij van mening dat toetsen op school noodzakelijke prikkels zijn die leerlingen aanzetten tot leren. Vervolgens ziet hij de behaalde resultaten op een toets ook als uitgangspunt voor een evaluatie van het onderwijs dat hij heeft aangeboden en verzorgd. Tot slot geeft hij aan dat positieve toetsresultaten een leerling stimuleren en motiveren.

Pete

Pete verwijst naar drie formatieve ervaringen in zijn onderwijs- en toetsingspraktijk: de ideeën en intensieve samenwerking met twee collega's op zijn eerste school die vonden dat leergangen nooit bieden wat je als docent wilt bereiken, zijn tweejarige ervaring als leerkracht op een andere school, waar *Pete* was aangetrokken om de didactiek te moderniseren van de sectie Engels, en, tot slot, het jarenlange vertrouwen en de autonomie die hij van zijn huidige schoolleiding altijd heeft gekregen.

Pete noemt de twee collega's die hem opvangen en begeleiden in zijn eerste baan progressief. Zij zijn actief betrokken bij onderwijsvernieuwing en nemen met regelmaat deel aan nacholingscursussen. De twee collega's geven *Pete* de kans zelf te ontdekken dat tekst- en werkboeken niet bieden wat je als docent aan je leerlingen kwijt wilt. Op zijn tweede school krijgt *Pete* te maken met een onwillige 5-havo klas. De leerlingen zijn niet bereid tijd en energie te steken in de taken die hij voor hen heeft bedacht. De leerlingen zijn gewend dat de leerkracht het werk voor hen doet. *Pete* lost het aanvankelijke conflict met de klas op en weet de leerlingen te overtuigen van het nut van zijn didactische aanpak. Als laatste formatieve ervaring noemt hij de kans die hem en zijn collega's wordt geboden door zijn schoolleiding om uit te voeren waar hij het meest in gelooft, dat wil zeggen lesgeven in de bovenbouw door middel van motiverende en zelfontwikkelde projecten, met veel authentieke taalinput en betekenisvolle taken. Ook *Pete's* formatieve ervaringen klinken door in zijn kernovertuigingen.

Pete gelooft in:

- het ontwikkelen van eigen lesmateriaal dat is gebaseerd op authentieke teksten; dit materiaal biedt volgens *Pete* zoveel meer dan de gebruikelijke leergangen Engels;
- collegiale samenwerking, vanaf de start van zijn loopbaan;
- professionele ontwikkeling die bestaat uit het kritisch beschouwen van bestaande praktijken en werkwijzen;
- de hoge mate van autonomie die docenten nodig hebben om uit te voeren waar zij in geloven;
- projectonderwijs waarin veel kennis en vaardigheden zijn geïntegreerd;
- het succes van zijn projectonderwijs, gezien de hoge scores die zijn leerlingen behalen op de nationale lees- en luistertoetsen, zonder daar veel voor geoefend te hebben;
- de kritische kanttekeningen die kunnen worden geplaatst bij de door het CITO ontwikkelde leesvaardigheidsexamens en de kijk-, luister- en schrijftoetsen;
- het belang van leerlinggerichtheid, die inhoudt dat de docent rekening houdt met de wensen van de leerlingen en met hetgeen zij weten en kunnen;
- de stelling dat toetsresultaten erg belangrijk zijn voor de leerlingen, hoewel hij zelf niet echt overtuigd is van het belang en de waarde van toetsen en proefwerken;
- het belang van het effectief en systematisch voorbereiden van leerlingen op toekomstige toetstaken;
- de beperkte mogelijkheid om toetsresultaten in een objectief cijfer uit te drukken bij complexe vaardigheden zoals schrijfvaardigheid;
- een holistische beoordeling van complexe vaardigheden, zonder harde beoordelingscriteria die tevoren zijn vastgesteld;
- integratieve toetsen, gebaseerd op de betekenisvolle input van aantrekkelijke teksten of beeldmateriaal;
- toetsen die meer vragen dan louter reproductie van uit het hoofd geleerde grammatica of vocabulaire zonder enige communicatieve context;
- het zorgvuldig plannen van oefentaken, activiteiten en toetsen in de loop van een schooljaar;

- de beperkingen van het nastreven van een zogenaamd “native-speaker” niveau, waarbij *Pete* de vraag stelt wat voor een niveau dan wel precies wordt bedoeld;
- het motiveren van leerlingen voor literaire teksten op een ontspannen manier, gebruikmakend van diverse media, in tegenstelling tot de vaak zwaarwichtige wijze waarop docenten literatuur en literaire teksten presenteren;
- de opvatting dat het lezen van een Engelse roman voor de meeste leerlingen een worsteling is;
- de noodzaak de onderbouw en bovenbouw beter op elkaar te laten aansluiten voor wat betreft de aangeleerde kennis en vaardigheden;
- het positief en realistisch benaderen van de vernieuwde Tweede Fase, omdat hij verwacht dat zijn didactiek goed past binnen de hervormingen.

In de bespreking van de autonomie van de leerling geeft *Pete* allereerst aan dat de groei naar meer autonomie geleidelijk moet verlopen. Leerlingen zijn niet gewend veel initiatief te nemen en verantwoordelijkheid te dragen voor hun eigen leren. Pete maakt zijn leerlingen zelfstandiger door zijn didactiek, die erop is gericht de moelijkheidsgraad van leer- en toetstaken geleidelijk te vergroten, aandacht te besteden aan de motivatie van leerlingen om te leren en vormen van samenwerkend leren in te bouwen.

Pete interpreteert communicatief taalonderwijs als onderwijs dat zich richt op het overbrengen van een boodschap in het Engels, waarbij de ontvanger bereid is de boodschap te begrijpen. Deze boodschap moet voor beiden betekenisvol zijn.

Een effectieve schriftelijke taaltoets Engels is volgens *Pete* communicatief, betekenisvol, gerelateerd aan wat onderwezen en geleerd is en gericht op transfer.

Hij onderscheidt vijf soorten essentiële kennis van en/of vaardigheden in het Engels die de leerlingen zich eigen moeten maken. Als eerste element noemt hij de kennis dat het onderwerp altijd voor het hoofdwerkwoord staat in het Engels. Vervolgens noemt hij de vaardigheid uit te drukken wat je in het Engels uit wil drukken, zonder daarbij teveel te leunen op directe vertaling van de moedertaal in het Engels. Ten derde noemt hij het inzicht van de leerling in hoeverre een gebruiker van het Engels de gesproken of geschreven taaluitingen van de leerling begrijpt. Als vierde element wijst *Pete* op het belang van kennis en effectieve toepassing van taalfuncties, zoals het vertellen hoe iemand op de plaats van bestemming komt. Een laatste essentieel aspect is de vaardigheid een eigen mening in het Engels te formuleren, bijvoorbeeld over een roman die de leerling heeft gelezen. Om dat te bereiken moet een leerling volgens *Pete* kunnen beschikken over een zekere woordenschat, een bepaald grammaticaal inzicht, kennis van de persoon tot wie men zich richt en kennis van de registers van taal, zoals formeel en informeel taalgebruik.

Met het oog op de potentiële terugslageffecten van toetsen merkt *Pete* op dat toetsen nooit als ultiem doel moeten worden gezien, maar ondergeschikt zijn aan het proces van almaar taalvaardiger worden. Pete geeft wel aan dat cijfers de leerlingen zowel kunnen aanmoedigen als ontmoedigen. Hij voegt tevens toe dat een teleurstellend toetsresultaat een leerling zou kunnen aanzetten tot nadenken over wat er precies is misgegaan.

De hoofdstukken **acht,** (*Joy, the budding professional*) ***negen*** (*Mark, the literary master*) en **tien** (*Pete, the project man*) richten zich op gegevens uit de drie vervolginterviews en de door de respondenten geselecteerde toetsen in de loop van het schooljaar. De hoofdstukken starten met een overzicht van de *core beliefs* en interpretaties van de drie centrale constructen, zoals die naar voren zijn gekomen in de eerste interviews. Daarna worden de drie geselecteerde toetsen besproken.

Joy selecteert als eerste toets een schrijftoets waarmee een project over tienerzwangerschappen wordt afgesloten. Het project is voor een groot deel

gebaseerd op Berlie Doherty's roman *Dear Nobody* en de verfilming van Roddy Doyle's *The Snapper*. De tweede toets die wordt besproken, is een grammaticatoets die kenmerkend is voor de wijze waarop zij grammatica wenst te toetsen. De toets is een bewerking van een tekst genomen uit de *Daily Mail*, met als titel *Little boy growing old before his time*. Als derde toets kiest *Joy* voor een leesvaardigheidstoets, waarmee een project over de Ierse kwestie wordt afgerond.

Mark verkiest als eerste toets een proefwerk over een kort verhaal *A Day's Wait* van Ernest Hemingway. Hij rondt met deze toets een serie lessen af die zijn gericht op het aanleren van begrippen die nodig zijn om literatuur in het Engels te bespreken en het opdoen van leeservaring met korte verhalen. Als tweede toets kiest *Mark* voor een reguliere toets over een hoofdstuk uit de leergang die wordt gebruikt. De derde toets van *Mark* is wederom een literaire toets. Voortbordurend op de literaire begrippen die de leerlingen hebben leren gebruiken en de leeservaringen die zij hebben opgedaan, krijgen kleine groepen leerlingen nu de opdracht een gekozen roman zelf te lezen, in de groep te bespreken en gezamenlijk te analyseren aan de hand van een geschreven opdracht en beoordelingscriteria die tevoren zijn verstrekt.

Pete geeft aan dezelfde toetsen te willen bespreken als collega *Joy*, enerzijds omdat hij het prettig vindt ervaringen met *Joy* uit te wisselen en anderzijds omdat hij het overwegend eens is met haar keuze.

Over iedere toets wordt op dezelfde wijze gerapporteerd. Allereerst geeft de respondent de *redenen* aan waarom de toets ter bespreking is gekozen. Daarna wordt ingegaan op de *specifieke kennis, vaardigheden en inzicht* die de toets volgens de respondent meet. Dan volgen bijzonderheden over *constructie en gebruik*, te weten hoe de toets tot stand is gekomen, hoe de docent aan de kennis en vaardigheden kwam om de toets te maken, de afnamecondities, de wijze waarop de toets is *gescoord, genormeerd en beoordeeld* en tot slot de wijze waarop de toetsresultaten zijn *geëvalueerd*. Tot slot rapporteren de hoofdstukken over eventuele *terugslageffecten* die de respondenten zien op de zelfstandigheid van hun leerlingen en de wijze waarop zij leren communiceren in het Engels.

De resultaten van de hoofdstukken 8, 9 en 10 worden aan het einde van ieder hoofdstuk in relatie gebracht met de kernovertuigingen van de respondenten uit hoofdstuk 7.

Bij *Joy* komen twaalf van de veertien *core beliefs* uit het eerste interview overtuigend terug in de selectie en bespreking van de drie toetsen in de loop van het schooljaar. Twee van haar opvattingen komen minder expliciet terug in de vervolginterviews. Met uitzondering van de *Dear Nobody* toets is geen overtuigend bewijs gevonden van *Joy's* opvatting dat leerlingen worden gemotiveerd als zij zelf bepaalde keuzes mogen maken. Met uitzondering van de *Little Boy* grammaticatoets is tevens geen overtuigend bewijs gevonden van haar voornemen om inzicht in hoe een taal werkt een steeds grotere rol te laten spelen in haar onderwijs- en toetsingspraktijk.

In het geval van *Mark* komen veertien van zijn twintig kernovertuigingen overtuigend terug in de selectie en bespreking van de drie gekozen toetsen. Van vier *core beliefs* is deels convergent bewijs gevonden. Het betreft zijn overtuigingen over een duidelijk onderscheid tussen de funderende onderbouw en de meer toepassingsgerichte bovenbouw, het overwegend negatieve beeld dat hij heeft over vernieuwingen van het onderwijs zoals de Tweede Fase, de tekortkomingen van de toetsen die bij de leergang worden geleverd en de opvatting dat literatuuronderwijs is gemarginaliseerd in de Tweede Fase. Twee kernovertuigingen komen niet meer expliciet terug in de loop van het schooljaar: de zinloosheid van de verplicht af te nemen Basisvormingstoetsen en de voordelen van leerlingen hun eigen fouten laten corrigeren.

Bij *Pete* is convergent bewijs gevonden van achttien van zijn twintig kernovertuigingen. Twee van zijn *core beliefs* kwamen niet meer expliciet terug in de vervolgdata, te weten het feit dat zijn didactische benadering de leerlingen goed

voorbereidt op de nationale CITO luister- en leestoetsen en de beperkingen van wat hij *native-speaker* niveau heeft genoemd.

In het **elfde** hoofdstuk, *Cross-case analyses*, worden de gegevens uit de hoofdstukken 7, 8, 9 & 10 gereduceerd, vergeleken en gecontrasteerd. De analyse is gericht op het vinden van overeenkomsten in de constructinterpretaties en de toetsingspraktijken van de respondenten, zonder daarbij de specifieke verschillen uit het oog te verliezen.

Ieder construct is in vier delen besproken: de interpretaties van het construct door de docenten, overeenkomsten in de interpretaties, de mate van convergentie van de interpretaties in de toetsingspraktijken van de informanten en een overzicht van wat is gevonden en wat niet is gevonden in relatie tot de inhoud van de theoretische hoofdstukken 3, 4 en 5 over *learner autonomy* (LA), *communicative language education* (CLE) en *foreign language assessment and evaluation* (toetsing, beoordeling en evaluatie).

Met betrekking tot het construct van de autonomie van de leerling zijn er vier overeenkomsten in de interpretaties van het begrip door de respondenten. Alleerst signaleren *Joy*, *Mark* en *Pete* dat de docent in eerste instantie het initiatief neemt, leertaken formuleert en vervolgens controle uitoefent op de voortgang. De docent stuurt de leerlingen in de richting van meer zelfstandigheid. Vervolgens hechten de informanten belang aan een geleidelijke overgang van docentgestuurd naar meer leerlinggestuurd onderwijs. Ten derde vinden *Joy, Mark* en *Pete* dat volledig zelfverantwoordelijk leren geen haalbaar begrip is in het voortgezet onderwijs. Tot slot vinden de respondenten het belangrijk dat leerlingen leren door te doen. Dat betekent dat leerlingen taken uitvoeren waarin in het Engels wordt gelezen, naar Engelse teksten gekeken en geluisterd en in de doeltaal moet worden gesproken en geschreven.

De visies op de sturende rol van de docent bij de ontwikkeling van zelfstandigheid van de leerling convergeerden het meest in de onderwijs- en toetsingspraktijken van de respondenten. Bewijs van de specifieke interpretaties van autonomie van de drie respondenten was minder overtuigend en vaak slechts impliciet aanwezig in hun onderwijs- en toetsingpraktijk. Zo was er slechts indirect bewijs van het belang dat *Joy* hecht aan het leerlingen laten reflecteren op wat zij doen, wat zij moeten doen, wat zij willen doen en wat daarvoor vervolgens nodig is. Hetzelfde geldt voor de interpretaties van *Mark*. De plannings-, zelfbeoordelings- en zelfevaluatievaardigheden en de vaardigheid om doe- en denkactiviteiten beredeneerd met elkaar af te wisselen, komen niet expliciet terug in de vervolgdata. Er is meer expliciete convergentie in het geval van Pete, met name voor het belang dat hij hecht aan het begrijpen en motiveren van zijn leerlingen.

Een vergelijking met de parameters van autonomie uit hoofdstuk 3, leidt tot tien overeenkomsten en zes aspecten die ontbreken in de onderwijs- en toetsingspraktijken van de respondenten. De overeenkomsten zijn het belang dat de docenten hechten aan het motiveren van hun leerlingen; het creëren van docentgestuurde leeromgevingen waar leerlingen uit kunnen drukken wie zij zijn en wat zij vinden; docentinterventies wanneer de respondenten dat noodzakelijk achten; het feit dat autonoom gedrag noch eenvoudig, noch objectief beschreven kan worden; erkenning dat autonomie alleen in relatieve termen kan worden besproken; bewijs dat een klein aantal leerlingen – hoofdzakelijk afkomstig uit de vwo-klassen - autonoom is als leerders, inzicht heeft en goed is in Engels zonder dat daar een duidelijke verklaring voor is; het belang van de kennis en vaardigheden die vereist zijn om leertaken succesvol uit te voeren alsook van het soort taak en de activiteiten die leerlingen uit moeten voeren en tot slot de opvatting dat de voorgestelde vernieuwingen van de Tweede Fase niet fundamenteel van invloed zijn op het didactisch denken en handelen van de drie informanten.

Er zijn ook zes aspecten die ontbreken: geen van de drie respondenten verwijst expliciet naar enige academische theorievorming op het gebied van de autonomie van de leerling; de overgrote meerderheid van de leerlingen kan in de beleving van de informanten niet zonder de sturing en controle van de docent in de groei naar meer autonomie; voorbeelden van leerlingen die reflecteren op wat zij leren, waarom zij leren, hoe zij leren en met welke mate van succes dat vervolgens gebeurt; structurele pogingen van de respondenten om de leerlingen hun eigen leerdoelen te laten bepalen, de inhoud en voortgang te bewaken, te kiezen uit een aantal nuttige leerstrategieën en de leerlingen een stem te geven bij initiatieven en besluiten en, tot slot, structurele pogingen om de leerlingen het eigen leren te laten plannen en de voortgang daarvan te laten beoordelen en te evalueren, zowel formatief als summatief.

Voor wat betreft het construct communicatief taalonderwijs zijn er twee overeenkomsten in de interpretaties van de respondenten. Allereerst blijken communicatief taalonderwijs en communicatieve competentie begrippen te zijn die zich lastig laten definiëren en specificeren. Ten tweede verwijzen alle respondenten in meer of mindere mate naar aspecten van vorm, gebruik en betekenis.

De interpretaties van het construct communicatief taalonderwijs convergeren in hoge mate in de onderwijs- en toetsingspraktijken van *Joy* en *Pete*. Bij *Mark* is geen bewijs gevonden van taalbeschouwing en het bewust maken van leerlingen voor contrastieve vormen en gebruik, met uitzondering van de relatief grote hoeveelheid vertaalitems in de Unicom-Finals toets.

Een vergelijking met de definities en kenmerken van communicatief taalonderwijs en communicatieve competentie uit hoofdstuk 4 levert tien overeenkomsten op en zeven aspecten die ontbreken. De overeenkomsten zijn dat de informanten taal zien als communicatie; impliciet vinden dat communicatie draait om het uitdrukken en interpreteren van wat betekenisvol is; details geven over hoe essentiële kennis en vaardigheden worden overgedragen van de docent op de leerling; conflicterende meningen hebben over het nut en het gebuik van leergangen; niet aantonen dat een bepaalde benadering of didactiek effectiever is in het leren communiceren van leerlingen dan een andere; kenmerken van de grammatica-vertaal-, de directe en de audiolinguale methode laten voorkomen in hun onderwijspraktijk; slechts twee perspectieven op communicatief taalonderwijs hebben; expliciete aandacht hebben voor grammaticale competentie binnen de linguistische competentie; enige aandacht hebben voor discourse-competentie en het frequent gebruik van literaire teksten.

Zes aspecten ontbreken in de interpretaties van de drie informanten van het begrip communicatief taalonderwijs. Alleereerst verwijzen de drie respondenten niet naar enige academische theorievorming. De informanten gaan in hun interpretaties niet in op de specificering van de leercontext en formuleren geen heldere communicatieve doelen, met aandacht voor de kennis en vaardigheden die nodig zijn om deze doelen te realiseren en de functies, noties en strategieën die daarbij centraal staan. *Joy, Mark* en *Pete* zijn niet expliciet gericht op het belang van het onderhandelen van betekenis in communicatieve situaties. Zij richten zich nauwelijks op lexicale, semantische en fonologische competenties als onderdelen van linguistische competentie. Daarnaast ontbreekt grotendeels de aandacht voor sociolinguistische, pragmatische en strategische competenties. Er zijn slechts beperkte mogelijkheden voor leerlingen om hun Engels te oefenen en constructieve feedback te krijgen op hoe zij oefenen, hoe zij leren en de mate van succes waarmee dat gebeurt. Tot slot, expliciete pogingen om het bewustzijn van leerlingen te vergroten over vorm, gebruik en betekenis als zij luisteren naar, lezen, spreken of schrijven in het Engels ontbreken grotendeels.

De bespreking van wat de informanten zien als effectieve toetsing valt uiteen in drie aspecten: wat zij zien als een effectieve schriftelijke toets Engels, de essentiële kennis van en vaardigheden en inzicht in het Engels die zij bij hun leerlingen willen

toetsen en de mogelijke terugslageffecten van toetsen op de manier waarop leerlingen leren communiceren in het Engels.

Er zijn twee overeenkomsten in wat de informanten zien als een effectieve schriftelijke toets Engels. Allereerst hechten zij belang aan met name de inhoudsvaliditeit, zonder het concept als zodanig te benoemen. Ten tweede vinden zij dat een toets hoofdzakelijk toepassingsgericht moet zijn, d.w.z mogelijkheden biedt tot transfer van wat is geleerd.

Er is een hoge mate van convergentie voor de percepties van inhoudsvalideit in de toetsingspraktijk van de informanten. Bij *Joy* wordt echter geen bewijs gevonden dat een toets iets moeilijker moet zijn dan de taken waarmee is geoefend. *Marks* interpretatie dat leerlingen in een effectieve toets de grammatica en het bestudeerde idioom zelf creatief in zinnen moeten gebruiken, convergeert niet in de reguliere idioom-, grammatica- en leesvaardigheidstoetsen die de leerlingen met regelmaat krijgen. De creatieve toepassing betreft hier louter vertaaloefeningen, waarbij reproductie een belangrijke rol speelt. Die creatieve toepassing speelt een grotere rol bij de literaire toetsen en taken die de leerlingen maken. *Pete's* opvattingen over een effectieve taaltoets convergeren volledig in zijn onderwijs- en toetsingspraktijk.

Er zijn twee overeenkomsten in de specifieke kennis en vaardigheden die het inzicht dat de informanten belangrijk vinden om te toetsen. Allereerst blijkt het problematisch om kennis, vaardigheden en inzicht duidelijk van elkaar te onderscheiden. Ten tweede hebben de genoemde kennis, vaardigheden en inzicht betrekking op de linguïstische competentie, met de nadruk op grammaticale competentie. Dat is dan ook de reden dat het bewijs met name in de meer op grammatica gerichte toetsen convergeert. Linguïstische kennis, inzicht en vaardigheden spelen slechts een indirecte rol in de schrijf- en literaire toetsen.

Er zijn twee overeenkomsten in de percepties van de informanten van mogelijke terugslageffecten van toetsen op het leren communiceren in het Engels. Allereerst geven zij allen aan dat de leerlingen cijfers belangrijk vinden. Als tweede overeenkomst noemen zij dat cijfers zowel kunnen motiveren als demotiveren. Toch convergeren de in potentie positieve terugslageffecten van toetsen niet in de toetsingspraktijk van de informanten. *Joy* laat als enige docent leerlingen vaak zelf de gemaakte fouten corrigeren en bespreekt de gemaakte toetsen veelal consequent na, maar zij vindt dat leerlingen onvoldoende reflecteren op de fouten die zij hebben gemaakt. *Mark* en *Pete* lijken de potentiële terugslageffecten van hun toetsen te laten voor wat zij zijn.

Een vergelijking met de inhoud van het theoretische hoofdstuk over toetsing, beoordeling en evaluatie toont allereerst aan dat de drie informanten toetsen zien als summatieve beoordelingsmomenten. Hoewel de oefentaken richting eindtoets zorgvuldig zijn opgebouwd, is hierbij niet of nauwelijks sprake van formatieve beoordeling en evaluatie. Daarnaast blijkt dat de informanten met regelmaat toetsen en beoordelen, maar zelden expliciet en of succesvol toekomen aan het evalueren van de afgenomen toetsen met hun leerlingen. Verder blijkt dat de *Joy*, *Mark* en *Pete* vooral zijn gericht op prestatietoetsen (*achievement tests*) en niet op vaardigheidstoetsen (*proficiency tests*).

Van de vier besproken trends op het terrein van toetsing, komen een aantal kenmerken van de voorwetenschappelijke, de psychometrisch-structuralistische en de integratief-sociolinguistische trend terug in de toetsingspratijken. Dit geldt echter niet voor de meer recente aandacht voor de kritisch-dynamische trend. Alle informanten lijken een voorkeur te hebben voor communicatieve schrijftoetsen, al dan niet op literaire teksten of literaire en/of culturele kennis gebaseerd. Tot slot benoemen de informanten impliciet en niet al te gedetailleerd het belang van de toetskwaliteiten betrouwbaarheid en validiteit.

Het **twaalfde** hoofdstuk, *Autonomy tested: A discussion,* bestaat uit vijf delen. In het eerste deel wordt de inhoud van voorgaande hoofdstukken teruggehaald en

worden de onderzoeksvragen nogmaals gepresenteerd. Er wordt vastgesteld dat vier van de zeven deelvragen van het onderzoek al zijn beantwoord in voorgaande hoofdstukken. Het slothoofdstuk richt zich vervolgens op het beantwoorden van de drie resterende deelvragen.

In het tweede deel wordt de vraag beantwoord in hoeverre de onderzoeksresultaten overeenkomen met de theoretische verkenningen van de drie centrale begrippen in de studie. De drie constructen worden op dezelfde wijze besproken. Eerst wordt de kerntheorie besproken uit het betreffende theoretische hoofdstuk. Vervolgens worden potentiële overeenkomsten genoemd tussen de kerntheorie en de interpretaties van de respondenten. Daarna worden de te overbruggen hiaten besproken tussen de interpretaties van de respondenten en de theorie. Tot slot volgt een analyse waarom de hiaten zijn zoals gevonden.

De autonomie van de leerling

Vier thema's uit hoofdstuk 3 zijn in het bijzonder van nut bij de analyse en bespreking van de interpretaties van de drie informanten. Allereerst zijn dat de dertien parameters van autonomie die zijn geïdentificeerd op basis van een verkenning van gangbare definities van het construct *learner autonomy*. Ten tweede zijn dat de *reasons-responsive* en *responsiveness-to-reasoning* interpretaties van autonomie binnen de filosofie. Beide interpretaties stimuleren in potentie de metacognitieve en metalinguistische reflectie van leerlingen als zij leren communiceren in een andere taal. Ten derde helpen humanistische, cognitieve, constructivistische en sociaal-culturele perspectieven op menselijk leren het denken en handelen van leerlingen te begrijpen en theoretisch te verklaren. Het vierde belangrijke thema uit hoofdstuk 3 is kennis van motivatietheorieën, gezien de symbiotische relatie van motivatie en autonomie in onderwijsleersituaties.

Er zijn drie overeenkomsten tussen de kernovertuigingen van de informanten over de autonomie van de leerling en de besproken theorie. *Joy*, *Mark* en *Pete* zien autonomie als een relevant pedagogisch doel. Hierbij is een belangrijke rol weggelegd voor de sturende leraar. Een tweede overeenkomst is het belang dat de informanten hechten aan het welzijn, de motivatie en het motiveren van hun leerlingen. Ten derde presenteren de informanten interessante en elkaar aanvullende interpretaties van autonomie.

Er zijn drie te overbruggen hiaten tussen de interpraties van autonomie en de besproken kernthema's in hoofdstuk 3. Allereerst blijkt dat de interpretaties van de docenten niet of slechts marginaal zijn gebaseerd op wat theoretisch over het construct bekend is. Er is een gat tussen de gepresenteerde praktijktheorieën en de besproken academische theorie dat gedicht dient te worden. Ten tweede blijkt het verre van simpel voor de informanten om hun interpretaties van autonomie te realiseren in de onderwijs- en toetsingspraktijk. Tot slot lijkt slechts een beperkt aantal leerlingen verantwoordelijkheid te nemen voor hun eigen leren. De inhoud van hoofdstuk 3 kan docenten helpen de kennis van autonomie en het handelingsrepertoire te vergroten bij het stimuleren van zelfstandigheid van leerlingen.

Zes redenen worden genoemd waarom theorievorming over zelfstandig leren een marginale rol heeft gespeeld in de lespraktijk van de informanten. Ten eerste zijn de uitgangstheorieën van de Tweede Fase noch breed, noch diepgaand geweest. Autonomie is op het moment van invoering nog slechts sporadisch aan de specifieke sociaal-culturele context van het Nederlandse onderwijs getoetst. Ten tweede zijn slechts enkele kernconcepten van de theorievorming rond de vernieuwing, zoals een "half-niet-centraal", zelfstandig werken, leerlingactivering, of planningsvaardigheden, in gefilterde vorm bij de informanten terecht gekomen. Een derde reden is dat bij de invoering van de vernieuwde Tweede Fase nog weinig empirisch bewijs is dat de innovaties ook daadwerkelijk leiden tot meer zelfstandigheid en een toename van kennis en vaardigheden van leerlingen. Een vierde reden is dat er bij de invoering

weinig tot geen tijd is voor docenten om over centrale concepten na te denken, daarmee te experimenteren en de resultaten te evalueren. Deze situatie verergert aanzienlijk als blijkt dat docenten les moeten geven aan meer en grotere klassen, met minder contacttijd. Een vijfde reden is de impliciete aanname dat de kernovertuigingen van docenten kunnen worden beïnvloed door vernieuwingen top-down in te voeren. Een zesde reden waarom theorie weinig aandacht krijgt, ligt bij de docenten zelf. Professionele ontwikkeling en reflectie op wat docenten denken en doen, op welke momenten dat gebeurt en waarom is geen vanzelfsprekendheid voor de meer ervaren informanten uit het onderzoek.

Bovenstaande redenen lijken te hebben beïnvloed waarom de informanten hun voornemens voor het bevorderen van de zelfstandigheid van leerlingen moeilijk kunnen realiseren in de lespraktijk. De druk is groot, voor zowel docenten als voor leerlingen, de tijd voor reflectie ontbreekt en praktische richtlijnen die helpen de zelfstandigheid van leerlingen te bevorderen zijn niet altijd voorhanden. Daar komt bij dat de informanten twijfelen aan de haalbaarheid van autonomie voor adolescente leerlingen, die veelal verplicht het regulier onderwijs volgen. De theorie ondersteunt de informanten in hun visie dat volledige zelfverantwoordelijkheid voor het eigen leren niet haalbaar is voor het gros van de leerlingen. Toch laat zowel theorie als empirie zien dat de ontwikkeling naar meer zelfstandigheid bij mensen plaatsvindt vanaf het moment dat zij worden geboren, vaak ondanks opvoeding en onderwijs.

Er zijn drie redenen waarom de autonomie van leerlingen als pedagogisch doel lastig haalbaar is voor de drie informanten. Allereerst laten de data van het onderzoek zien dat de docenten er niet in zijn geslaagd de leerlingen te laten begrijpen en uit te laten leggen wat zij leren, waarom zij leren, hoe zij leren en met welke mate van succes dat gebeurt. Ten tweede laat de studie zien dat de informanten zich slechts richten op een beperkt aantal van de parameters van autonomie die zijn gedefinïeerd. Belangrijke parameters, zoals het (mede)bepalen van leerdoelen, de lesinhoud en de voortgang en het leerlingen laten kiezen uit belangrijke leer- en denkstrategieën, blijven grotendeels buiten beschouwing. Tot slot zijn er weinig gerichte pogingen van de informanten om leerlingen het eigen leren doelmatig te laten plannen en zelf de specifieke voortgang te laten beoordelen en evalueren, zowel formatief als summatief. Veel van de parameters van autonomie zijn didactisch onontgonnen.

Communicatief taalonderwijs en communicatieve competentie

Drie thema's uit het theoretisch hoofdstuk over communicatief taalonderwijs zijn in het bijzonder van belang. Allereerst is dat de kennis van de methodologische geschiedenis van het leren en onderwijzen van andere talen. Ten tweede is dat Savignon's definitie van communicatie als het voortdurende proces van het uitdrukken, interpreteren en onderhandelen van betekenis en zijn dat de modellen van communicatieve competentie die laten zien dat het construct uit meerdere kennis- en vaardigheidsgebieden bestaat. Een derde relevant thema is de vraag hoe aandacht voor vormaspecten van taal, zoals uitspraak, spelling, het gebruik van de juiste vorm en tijd of de receptieve en productieve woordenschat waarover een leerder beschikt, zich verhoudt tot de betekenis die in geschreven en gesproken taal wordt uitgedrukt.

Er worden twee overeenkomsten genoemd tussen de interpretaties van de informanten en de beproken theorie. Alleerst geloven de drie docenten in taal als middel tot communicatie en geloven zij impliciet in communicatie als het uitdrukken en interpreteren van bepaalde betekenis. Ten tweede erkennen de drie informanten, ieder op zijn of haar eigen wijze, het belang van authenticiteit. Er moet daadwerkelijk iets te communiceren zijn in het Engels, wat niet volledig van tevoren voorspeld kan worden en waarin de leerling zichzelf laat kennen als persoon.

Er zijn ook vier hiaten tussen wat de informanten denken, voelen en doen en de besproken theorie. Allereerst is dat wederom het gemis tussen wat bekend is over

het construct en de praktijktheorieën van de informanten. Dit leidt tot een beperkte opvatting van communicatie en communicatief taalonderwijs. Deze beperkte opvatting is opmerkelijk gezien de aandacht in Nederland voor communicatief taalonderwijs en communicatieve leergangen sinds de jaren tachtig van de vorige eeuw. De beschikbare tijd kan hier moeilijk als reden worden genoemd. Evenmin kan de druk van de Tweede Fase deze beperkte visie hebben veroorzaakt. Feit is dat slechts een beperkt aantal aspecten van communicatieve competentie onderdeel is gaan uitmaken van de praktijktheorieën van de informanten. De casestudy's van dit onderzoek laten zien dat het gebruik van meer communicatieve lesmaterialen en lestaken niet automatisch leidt tot leerlingen die beter hebben leren communiceren in het Engels. Communicatieve competentie is een construct dat tenminste zo complex en gelaagd is als de autonomie van de leerling. Er wordt beargumenteerd dat meer academische kennis de docenten kan helpen bij wat grotendeels in de data ontbreekt.

Als tweede hiaat wordt genoemd het gebrek aan specificering van wat communicatief taalonderwijs en communicatieve competentie nu precies is. Kennis van en ervaringen met definities van communicatie en modellen van communicatieve competentie en de mogelijke operationalisaties daarvan, kunnen docenten helpen de leerlingen vaardiger te maken als gebruikers van het Engels.

Een derde hiaat is de beperkte aandacht voor de ontwikkeling van het taalbewustzijn van leerlingen van de relaties tussen vorm, gebruik en betekenis. Bij de bespreking van het construct autonomie is al vermeld dat de informanten zich slechts in beperkte mate richten op metacognitieve en metalinguïstische reflectie en op formatieve beoordeling en evaluatie. Reflectie en evaluatie worden gezien als belangrijke componenten. Het is van belang leerlingen bewust te maken van de relatie tussen de vormaspecten van taal, de specifieke gebruiksaspecten en de uiteindelijke betekenisverlening die wordt nagestreefd.

Tot slot is er weinig tot geen aandacht voor het onderhandelen van betekenis wanneer twee of meer personen communiceren. Er wordt beargumenteerd dat als de schrijf- en literaire toetsen uit dit onderzoek niet worden geëvalueerd met en door de leerlingen, zij geneigd zijn eerder verworven kennis en ideeën simpelweg te reproduceren. Duidelijke voorbeelden van het onderhandelen van de betekenis van wat leerlingen hebben geschreven zijn niet gevonden in het onderzoek.

Er zijn twee bevindingen die niet rechtstreeks aan de inhoud van het theoretische hoofdstuk over communicatief taalonderwijs en communicatieve competentie kunnen worden gekoppeld. Het betreft de rol van de gebruikte leergangen en het gebruik van literaire teksten wanneer leerlingen leren communiceren in het Engels.

De gebruikte leergang is belangrijk voor *Mark*. De selectie en gradatie van wat in de leergang wordt aangeboden, bepaalt in hoge mate welke kennis en welke vaardigheden worden aangeboden en hoe in de leergang communicatief taalonderwijs wordt geïnterpreteerd. Het gebruik van de leergang leidt bij Mark niet tot spectaculaire veranderingen in waar hij al jaren in gelooft. De selectie en gradatie van leerstof komt in hoge mate overeen met zijn kernovertuigingen. *Joy* en *Pete* zijn het schooljaar gestart met het gebruik van een leergang, maar concluderen al snel dat de communicatieve doelen die zij willen bereiken beter behaald kunnen worden met behulp van motiverende projecten, die zij zelf ontwerpen en waarin een aantal communicatieve taken zijn geïntegreerd. De conclusie hierbij is dat de leergangen slechts beperkt van invloed zijn geweest op de kernovertuigingen van de docenten en dat zij er niet wezenlijk toe hebben bijgedragen het onderwijs communicatiever te maken. Meer onderzoek is gewenst naar de specifieke rol van gebruikte lesmaterialen en aanbevolen didactieken als *agents of change*. Vooralsnog lijkt de mate waarin de leergang aansluit bij de overtuigingen van docenten belangrijk voor efficiënt en succesvol gebruik.

Alle informanten maken in meer of mindere mate gebruik van literaire teksten in hun oefen- en toetstaken. Dit gegeven suggereert dat gedichten, song- of rapteksten,

korte verhalen, toneelstukken, filmscripts en romans de leerlingen kunnen uitdagen tot betekenisvolle communicatie in het Engels. Dit wekt geen verbazing, omdat literaire teksten vaak voorbeelden zijn van het best mogelijke taalgebruik over thema's die als relevant en aansprekend worden gezien. Regelmatig gebruik van gevarieerde literaire teksten wordt dan ook gepropageerd als leerlingen leren communiceren in het Engels. Dit contrasteert met de vaak verheven en weinig concrete literaire doelen die in de Tweede Fase voor het literatuuronderwijs zijn gesteld, waarbij de inhoud van teksten vaak in het Nederlands wordt besproken en getoetst. Er wordt te gemakkelijk aangenomen dat het lezen van een kort verhaal of een roman in een andere taal de leerling te weinig interesseert en te veel energie kost. Het succesvolle gebruik van literaire teksten in het onderwijs van *Joy, Mark* en *Pete* laat zien dat de literaire tekst een plaats verdient in het vaardigheidsonderwijs, waarbij leerlingen keuzes worden geboden en leeservaringen opdoen, die zij vervolgens in de doeltaal bespreken met elkaar en met de docent. Literatuur lezen is deels leren over het leven en over jezelf. In die hoedanigheid speelt de literaire tekst een belangrijke rol bij zowel het ontwikkelen van taalvaardigheid als de ontwikkeling van meer kennis, begrip en autonomie.

Het toetsen, beoordelen en evalueren van leerprestaties in een andere taal

Vijf thema's uit hoofdstuk 5 hebben met name geholpen de gegevens van de drie informanten over toetsing, beoordeling en evaluatie te kunnen plaatsen. Allereerst is dat het gemaakte onderscheid tussen toetsing, beoordeling en evaluatie. Dit onderscheid wordt niet altijd helder gemaakt in zowel de academische theorie als in de praktijktheorieën van de drie docenten. Het belang van (zelf)evaluatie is daarbij benadrukt. (Zelf)evaluatie wordt gezien als een belangrijke verbinding tussen de constructen autonomie en communicatief taalonderwijs en communicatieve competentie, vanwege de gerichtheid op definieerbare en observeerbare kennis of vaardigheden en de relatie van evaluatie met metacognitieve en metalinguistische reflectie. Een tweede thema uit hoofdstuk 5 is de bespreking van de vier eerdergenoemde trends op het gebied van het toetsen van een andere taal. Een derde relevant thema is de relatie tussen de specifieke doelen van een toets en het soort toets. Een optimale relatie tussen toetsdoel en toetssoort helpt de validiteit van beoordelingsprocedures te vergroten. Een vierde thema is de aandacht voor de toetskwaliteiten van betrouwbaarheid en validiteit. Deze toetskwaliteiten zijn onderdeel van een model van zes kenmerken van een effectieve toets door Bachman & Palmer (1996). Dit model is het zesde relevante thema uit hoofdstuk 5.

Er zijn twee overeenkomsten tussen de praktijktheorieën van de informanten en de besproken thema's in hoofdstuk 5. Allereerst is er de erkenning dat cijfers belangrijk zijn voor leerlingen. Hiermee erkennen de informanten impliciet het belang van terugslageffecten van toetsen op hoe leerlingen leren en hoe docenten doceren. Een tweede overeenkomst bestaat uit de pogingen van de docenten de toetsen zo inhoudsvalide mogelijk te maken. De informanten vinden het belangrijk dat toetsen zowel leerbaar als op niveau zijn.

Er is ook sprake van zes hiaten. Het eerstgenoemde hiaat is wederom het feit dat academische theorievorming op het gebied van toetsing, beoordeling en evaluatie niet of nauwelijks van invloed is geweest op de praktijktheorieën van de drie informanten. Met name op het gebied van taaltoetsing is de beschikbare literatuur overweldigend. Wel vraagt het begrijpen van de literatuur specialistische kennis van de gebuikte vaktaal. Een tweede tekortkoming is dat de informanten weliswaar toetsen en beoordelen, maar dat zij de leerresultaten slechts zelden of met grote moeite evalueren met of laten evalueren door hun leerlingen. Hierdoor leren leerlingen veelal niet gestructureerd van de gemaakte fouten. Meer aandacht voor evaluatie is gewenst. De derde bevinding is dat de informanten toetsing hoofdzakelijk zien als summatieve beoordelingsmomenten aan het einde van een afgerond leerproces. Formatieve beoordeling en evaluatie zijn idealiter onderdeel van een

doordacht onderwijsleerprogramma. Een vierde bevinding is dat de toetsen die de informanten afnemen overwegend prestatietoetsen zijn. Er worden over het algemeen geen proficiency-toetsen afgenomen. Bij een proficiency-toets worden kennis en vaardigheden gemeten zonder dat de leerlingen voor de toets hebben geleerd. Vaak zijn dit soort toetsen ook gestandaardiseerd. Proficiency-toetsen zouden vaker moeten worden afgenomen en nabesproken in het voortgezet onderwijs. Ten vijfde blijkt het over het algemeen lastig voor de informanten aan te geven welke kennis of vaardigheden precies nodig zijn om de schrijf- of literaire toetsen met succes te maken. Als docenten dit al lastig vinden, dan heeft dit naar alle waarschijnlijkheid consequenties voor het leren van de leerlingen en werkt dat het maken van oefeningen op intuitie en gevoel in de hand. Tot slot zien we een paar kenmerken terug van de voorwetenschappelijke, psychometrisch-structuralistische en integratief-sociolinguïstische trends op het gebied van taaltoetsing, maar is de kritisch-dynamische trend grotendeels aan de aandacht van de informanten voorbijgegaan. Juist in deze laatste trend vindt men vernieuwingen, zoals aandacht voor *collaborative dialogue,* de wederzijdse verantwoordelijkheid voor het leerproces van de toetser en getoetste, de rechten van de getoetste en vormen van *dynamic assessment.* Deze vernieuwingen zijn gerelateerd aan interactionistische en interventionistische benaderingen van taaltoetsing en aan interne en externe vormen van formatieve beoordeling en evaluatie. Met name op dit terrein valt nog veel te leren en nog veel te onderzoeken.

Het derde deel van het slothoofdstuk richt zich op de vraag in hoeverre de vernieuwingen van de Tweede Fase hebben bijgedragen aan het vergroten van de autonomie van de leerlingen, het stimuleren van communicatieve vaardigheden in het Engels en het ontwikkelen van een effectieve beoordelings- en evaluatiepraktijk. De gegevens van de drie informanten laten zien dat de vernieuwingen niet aantoonbaar hebben geleid tot het stimuleren van communicatieve vaardigheden of het ontwikkelen van een effectieve beoordelings- en evaluatiepraktijk. Twee bescheiden successen op het gebied van de autonomie van de leerling komen naar voren. De vernieuwingen laten de docenten nadenken over de grenzen van autonomie. Zij komen tot interessante perspectieven op de zelfstandigheid van de leerling. Tevens wordt in de drie onderwijs- en toetsingspraktijken melding gemaakt van vormen van succesvol samenwerkend leren door leerlingen.

Zes redenen worden genoemd waarom de vernieuwde Tweede Fase slechts in beperkte mate heeft geleid tot verandering in het denken en handelen van de drie docenten. Een eerste reden is de aanname dat vernieuwingen in het onderwijs top-down kunnen worden geïntroduceerd, zonder daarbij rekening te houden met waar docenten in geloven en wat zij als haalbaar zien. Ten tweede richt de vernieuwing zich op zowel een ingrijpende innovatie van het curriculum als op wijzingen in de didactiek. Een volgende reden is dat de theoretische achtergrond van de Tweede Fase constructivistisch kan worden genoemd, maar dat aan die achtergrond weinig gerichte en empirische aandacht is besteed. Een vierde reden is dat docenten bij de invoering meer en over het algemeen grotere klassen krijgen. Dit wekt op zijn minst de schijn dat rendement en efficiency belangrijker zijn dan de kwaliteit van het onderwijs. Ten vijfde is er de impliciete aanname dat als leerlingen zelfstandig *werken*, zij ook daadwerkelijk zelfstandig *leren*. Die veronderstelde relatie is verre van vanzelfsprekend. Tot slot lijken het overladen curriculum en het gebrek aan gerichte begeleiding door docenten te veroorzaken dat leerlingen over het algemeen productgericht zijn. Het voltooien van het product lijkt belangrijker dan de kwaliteit van het leerproces.

Omdat bovenstaande bevindingen zijn gebaseerd op slechts drie lespraktijken en de gegevens niet zonder meer kunnen worden gegeneraliseerd, wordt ingegaan op twee studies die bovenstaande bevindingen lijken te bevestigen. Het gaat om het onderzoeksrapport *Zeven Jaar Tweede Fase: Een Balans* (Tweede Fase Loket, 2005) en een Europese studie genaamd *The Assessment of Pupils' Skills in English*

in Eight European Countries 2002 (Alabau, I., Bonnet, G., de Bot, K., Bramsby, J., Dauphin, L., Erickson, G., et al., 2004). Het onderzoeksrapport van het Tweede Fase Loket ondersteunt de bevindingen uit dit onderzoek dat meer aandacht moet worden besteed aan de zowel de communicatieve als de analytische vaardigheden als een moderne vreemde taal wordt onderwezen en geleerd. Het tweede onderzoeksrapport laat zien dat havisten en vwo-ers slechts gemiddeld scoren op hun kennis van en vaardigheden in het Engels in vergelijking met leerlingen elders in Europa.

Het vierde deel van het slothoofdstuk gaat in op de opbrengst voor vervolgonderzoek en voor cursussen gericht op positieve terugslageffecten van toetsing op hoe docenten onderwijzen en hoe leerlingen leren.

Gebaseerd op de resultaten van dit onderzoek, zijn er een viertal aanbevelingen: starten met waar docenten in geloven, het faciliteren van professionele ontwikkeling, het overbruggen van hiaten tussen theorie en praktijk en het pleiten voor reflectief actieonderzoek en (vak)didactisch ontwikkelingsonderzoek.

Starten met waar docenten in geloven

Uitgangspunt van zowel de professionele ontwikkeling van docenten als hervormingen van het onderwijs moet zijn wat docenten denken, vinden en voelen bij de start van een ontwikkelingstraject. Laat docenten uitdrukken, interpreteren en bespreken wat zij zien als effectief onderwijs, effectief leren en effectieve toetsing. Structurele dialogen met en tussen docenten blijken allerminst vanzelfsprekend, waardoor overtuigingen vaak verborgen blijven of slechts kunnen worden afgeleid door de les- en toetsingspraktijk te bestuderen.

Het faciliteren van professionele ontwikkeling

Docenten moeten in staat worden gesteld een beroepsleven lang te blijven leren. Daarvoor is allereerst tijd nodig. Reflectie en professionele ontwikkeling kosten tijd. Tijd is geld. Dat is met name het geval in kennismaatschappijen waar het accent ligt op efficiëntie, meritocratisch denken en individueel gewin in plaats van op collectieve ontwikkeling en gemeenschappelijk welzijn. Zo wordt er niet of niet op de juiste wijze geïnvesteerd in het onderwijs door politici en leidinggevenden. Hierbij lijkt het korte-termijn denken een lange-termijnvisie te domineren. Dit komt beoogde verbeteringen van de kwaliteit van het lesgeven en de kwaliteit van het leren van leerlingen absoluut niet ten goede.

Theorie en praktijk

De gegevens uit deze studie laten zien dat het belangrijk is dat docenten op de hoogte zijn van kernpublicaties over thema's, noties of vakdidactische concepten die gerelateerd zijn aan wat zij geloven, weten of doen.

In deze studie zijn voorbeelden besproken van kerntheorie of kernpublicaties voor de drie centrale constructen *autonomie van de leerling*, *communicatief taalonderwijs* en *beoordeling en evaluatie*. De analyse van de onderzoeksgegevens heeft aandachtsgebieden en thema's opgeleverd die kunnen aanzetten tot de professionele ontwikkeling van docenten en de verbetering van de kwaliteit van onderwijs.

Praktijkgericht actie- en ontwikkelingsonderzoek

De vierde en laatste aanbeveling van deze studie is een pleidooi voor vormen van onderzoek die docenten in staat te stellen hun lespraktijk systematisch te onderzoeken in samenwerking met anderen. Met Lewin (1948) zijn wij van mening dat een leraar tevens een onderzoeker moet zijn. Wij vinden twee praktijkgerichte onderzoeksparadigma's bij uitstek geschikt om bovenstaande aanbevelingen te realiseren. Het betreft *reflectief actie-onderzoek* (Curry, 2006; Flamini & Jiménez Raya, 2007) en *didactisch ontwikkelingsonderzoek* (Cobb, P., Confrey, J., deSessa,

A., Lehrer, R., & Schauble, L., 2003; Design-based research collective, 2003; Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006).

Het laatste deel van het slothoofdstuk gaat in op zeven tekortkomingen van de studie *Testing for Autonomy.* Allereerst wordt genoemd dat het aanvankelijke onderzoeksplan en de beoogde doelen en vragen te breed zijn geformuleerd en te ambitieus zijn. Vervolgens wordt vermeld dat de data aanvankelijk op een tijdrovende wijze zijn geanalyseerd. Een derde tekortkoming is het gebrek aan convergentie of het slechts deels convergeren van gegevens in de onderwijspraktijken van de informanten, wat niet met absolute zekerheid betekent dat de gegevens ook daadwerkelijk hebben ontbroken in het jaar van de dataverzameling. De contacttijd van de onderzoeker met de drie informanten en de klassen waaraan zij les gaven is daarvoor te beperkt geweest. Een vierde kritische kanttekening heeft te maken met de representativiteit van de informanten. Er kan niet worden geclaimd dat de drie docenten representatief zijn in positivistische zin en dat de gegevens kunnen worden gegeneraliseerd naar andere eerstegraads docenten Engels in Nederland. Een vijfde tekortkoming is het gebrek aan formele leerlinggegevens. Hoewel met leerlingen is gesproken en lessen zijn geobserveerd, hebben deze gegevens geen formele plaats gekregen in dit onderzoek. Een zesde aandachtspunt is de beperking van het onderzoek tot schriftelijke toetsen. Onderzoek naar communicatief taalonderwijs zou idealiter ook mondelinge toetsen en examens bij het onderzoek moeten betrekken. De laatste kanttekening is het dilemma waarvoor meer kwalitatieve onderzoekers zich gesteld zien. Het probleem is om op geloofwaardige wijze inzicht te geven in de stappen van het onderzoeksproces en dat vervolgens op economische wijze te doen. De gemaakte keuze voor exploratief onderzoek naar drie constructen en de presentatie van veel directe onderzoeksgegevens heeft niet alleen het proefschrift, maar ook deze samenvatting omvangrijk gemaakt.

CURRICULUM VITAE

Adri Elsen was born as the youngest son of a family of seven in Wamel, the Netherlands on 12th May1956. His parents soon moved to Nijmegen, where Adri attended primary school, secondary school and subsequently got his MA in English Language and Literature at the Catholic University of Nijmegen, with American Literature, General Literary Studies and Applied Lingustics as subsidiary subjects. His MA-thesis *Spiritual Conflict in 'The Temple'* was a comparative study of George Herbert's (1593-1633) religious poetry. His applied linguistics thesis *Proficiency level and the use of compensatory strategies by learners of English as a foreign language* was based on an investigation designed, carried out and reported on together with fellow students Marjon de Kleine, Pierre de Vries and Gerard Weijnen.

In the course of his studies, Adri and his student friends visited Great Britain a lot and were very much into sports, most notably general fitness, in- and outdoor soccer and basketball. In 1980 he was co-founder, first chairman and publicity hound of *Aeolus*, the student windsurfing club.

From 1976 till 1982, he always had jobs on the side. He subsequently worked as a salesman of electrical appliances, windsurfing gear and after getting his BA managed a photographer's shop for five months.

From September 1982 until October 1985 he worked as a student tutor of English at the Institute for Applied Linguistics (ITT) of his university. Here he developed a serious interest in the theory as well as practice of teaching and learning other languages. In 1983 he finished his degree-one teacher education course.

From November 1985 until 1st August 1986 he was a degree-one teacher of English in adult education at Dag- en Avondscholengemeenschap Craneveldt in Nijmegen. In addition he spent his Saturday mornings working as a teacher for Katholieke Leergangen Tilburg, a degree-two college for teacher education for half a year. From 1st August 1986 until 31st December 1997 he was a teacher of English in secondary education at Pax Christi College in Druten, initially part-time and later in full-time employment. From 24th November 1985 until 1st August 1991, he was a teacher educator at Pedagogische Hogeschool 'Oost-Gelderland' in Doetinchem, a teacher training college for primary school teachers. From 1991 until 1st January 1995, he was a school supervisor of student teachers of English, who were doing doing a degree-one teacher education course at Nijmegen university. On the 1st January 1995, he started his present profession as a teacher educator at UNILO, now the Graduate School of Education of Radboud University Nijmegen (ILS/RU). From 1998 onwards he has been a junior researcher for 1,5 days a week.

Together with his colleague Kees van Esch he joined two European projects. From 1998-2001 it concerned a Socrates Lingua-A project on learner autonomy in initial foreign language teacher education. From 2001-2003 it was a Comenius follow-up project that focused on teacher action research on the four language skills and on intercultural communicative competence. The results of the projects were published as volumes 8 and 9 of the Peter Lang series on *Foreign Language Teaching in Europe*.

When off work, Adri likes to read, sing, walk, run, cycle, sail, ride his motor bike and enjoy life to the full together with Carola, the love of his life.

ACKNOWLEDGEMENTS

Working on a dissertation for ten part-time years is no easy matter. Life happens to go on in the meantime. Life has first of all given me more and more students to teach. In addition, there have been serious health problems of people I love. Finally, my participation in projects and publications did not always help to speed things up. A lot of people kept me going, each in his/her own specific ways.

I could not possibly have finished this PhD. without the support of my wife, my dissertation supervisors, my relatives, my closest friends, my dear colleagues, and the friends I met in two three-year European research projects. I am indebted to each and everyone of them.

I would like to single out some names and especially thank them for their typical contribution to this thesis. If names happen to be missing, the mistake is all mine.

Ad Borsboom ; Alfons Andriessen ; Anita Engelen ; Astrid Eelkema ; Benno Elsen ; Carola de Haas ; Edith Verbeet ; Enrica Flamini ; Esther Schennink ; Eus Schalkwijk ; Fer Hooghuis ; Frans Arnold ; Gloria Jover ; Hans Fiere ; Harry Havekes ; Heleine Vedder ; Helga Deeg ; Helma Elsen-Janssen ; Helma Oolbekkink ; Hermine de Knijff ; Huub Hamers ; Jan de Vries ; Jan Sturm ; Jeanne Elsen-van Zijl ; Jeroen Imants ; Johan Luttenberg ; Jop van Amelsvoort ; Juan Espínola-Vazquez ; Jules Pieters ; Juul Willen ; Karin Derksen ; Kees de Bot ; Kees van Esch ; Klaas van Veen ; Laura Hermans ; Leon Vankan ; Lily van Wersch ; Lisette de Haas ; Lodewijk van Schalkwijk ; Manuel dos Santos Marieiro ; Manuel Jiménez-Raya ; Marcel Kamp; Marga Tesselhof-Elsen (†) ; Mariëlle Theunissen ; Marion Friedrich ; Marjon de Kleine ; Michiel Vogelezang ; Moeder Elsen ; Niels Brouwer ; Oliver St.John ; Paul Elsen ; Paul van Dijk ; Peter Sleegers ; Peter-Arno Coppen ; Pieter Kuipers ; Quincy Elvira ; Ragnar Aagård (†) ; Rainer Kaenders ; Simone Reis ; Steffie Klaassen ; Steven Uijen ; Theo Bergen ; Theo Marsman ; Thérèse Carpay ; Ton de Kok ; Ton Groos ; Wim Setz.